Access DB# 11669

SEARCH REQUEST FORM
Scientific and Technical Information Center

CRFE

Requester's Full Name: _McA Hines_   Examiner # : _____   Date: 5/3/01

Art Unit: _1045_   Phone Number 30 _5-0481_   Serial Number: _09426776_

Mail Box and Bldg/Room Location: _Anozez_ Results Format Preferred (circle): (PAPER) DISK E-MAIL

7E12

If more than one search is submitted, please prioritize searches in order of need.
****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _Isolated nucleic acids_

Inventors (please provide full names): _JEHK LING DING   DGUAN SOO TAN_
Bow HO

Earliest Priority Filing Date: _10/26/99_

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search SEQ ID NO 10 +
SEQ ID NO 11

Manu

nuc 11
prot 10

****

Conservation △

```
              GenCore version 4.5
         Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May  3, 2001, 15:51:41 ; Search time 17.84 Seconds
                                    (without alignments)
                                    137.969 Million cell updates/sec

Title:           US-09-426-776-10
Perfect score:   96
Sequence:        1 MRVLVLALAVAVGDGSNLG 21

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters:    374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL_15:*
      1:     sp_archea:*
      2:     sp_bacteria:*
      3:     sp_fungi:*
      4:     sp_human:*
      5:     sp_invertebrate:*
      6:     sp_mammal:*
      7:     sp_mhc:*
      8:     sp_organelle:*
      9:     sp_phage:*
     10:     sp_plant:*
     11:     sp_rodent:*
     12:     sp_unclassified:*
     13:     sp_vertebrate:*
     14:     sp_virus:*
```

                        SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	82	85.4	1788	13	Q9YGK0	Q9ygk0 oreochromis
2	48	50.0	330	10	Q9XIV8	Q9xiv8 nicotiana t
3	45	46.9	176	5	Q9W4U2	Q9w4u2 drosophila
4	45	46.9	262	10	Q9LV78	Q9lv78 arabidopsis
5	45	46.9	330	5	O18118	O18118 caenorhabdi
6	45	46.9	691	3	Q9UT43	Q9ut43 schizosacch
7	45	46.9	1718	5	Q9VGD4	Q9vgd4 drosophila
8	44	45.8	328	2	Q9RR76	Q9rr76 deinococcus
9	44	45.8	333	2	P96493	P96493 deinococcus
10	44	45.8	755	11	Q9R0G6	Q9r0g6 mus musculu
11	44	45.8	815	5	O44385	O44385 caenorhabdi
12	44	45.8	870	2	Q9WW45	Q9ww45 salmonella
13	44	45.8	992	2	O51918	O51918 peptostrept
14	44	45.8	1108	5	Q9U280	O9u280 caenorhabdi
15	44	45.8	1297	5	Q9V6P5	Q9v6p5 drosophila
16	43.5	43.3	952	2	O97198	O97198 leishmania
17	43	44.8	127	2	Q9RTH6	Q9rth6 deinococcus
18	43	44.8	181	4	Q9NTL2	Q9ntl2 homo sapien
19	43	44.8	273	2	O06170	O06170 mycobacteri
20	43	44.8	346	2	Q9PCH7	Q9pch7 xylella fas
21	43	44.8	356	5	Q9Y179	Q9y179 trichomonas
22	43	44.8	359	1	O47973	O47973 halobacteri
23	43	44.8	583	2	Q9RY54	Q9ry54 deinococcus
24	43	44.8	598	2	O83032	O83032 streptomyce
25	43	44.8	599	2	Q9L0I0	Q9l0i0 streptomyce
26	43	44.8	613	4	Q9NYM3	Q9nym3 homo sapien
27	43	44.8	729	5	Q9NTI2	Q9nti2 homo sapien
28	43	44.8	767	5	Q9V9Z6	Q9v9z6 drosophila
29	42.5	44.3	893	5	O16221	O16221 caenorhabdi
30	42.5	44.3	1113	2	Q91249	Q91249 streptomyce
31	42	43.8	85	5	Q9NLX8	Q9nlx8 leishmania
32	42	43.8	107	10	Q9LEP7	Q9lep7 brassica na
33	42	43.8	161	5	Q9U425	O9u425 manduca sex
34	42	43.8	174	1	O58538	O58538 methanococc
35	42	43.8	297	2	O29031	O29031 archaeoglob
36	42	43.8	338	2	Q9X595	Q9x595 rhodococcus
37	42	43.8	460	2	O06312	O06312 mycobacteri
38	42	43.8	634	2	Q9S6S8	Q9s6s8 rhizobium l
39	42	43.8	657	2	O86463	O86463 rhizobium l
40	42	43.8	1339	13	Q9W6I2	Q9w6i2 pimephales
41	41.5	43.2	296	5	Q9V7Z9	Q9v7z9 drosophila
42	41.5	43.2	351	2	Q9PL26	Q9pl26 chlamydia m
43	41.5	43.2	425	5	Q9VGH2	Q9vgh2 drosophila
44	41.5	43.2	1305	3	Q9P424	Q9p424 ajellomyces
45	41	42.7	109	2	Q9RWX6	Q9rwx6 deinococcus

                        ALIGNMENTS

```
RESULT   1
Q9YGK0
ID   Q9YGK0      PRELIMINARY;      PRT;   1788 AA.
AC   Q9YGK0;
DT   01-MAY-1999 (TREMBLrel. 10, Created)
DT   01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT   01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE   VITELLOGENIN PRECURSOR.
GN   VTG1.
OS   Oreochromis aureus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidei;
OC   Cichlidae; Oreochromis.
OX   NCBI_TaxID=47969;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Lim E.H., Lam T.J., Ding J.L.;
RT   "Cloning of Full-Length Oreochromis aureus vitellogenin cDNA and its
RT   structural and functional analyses.";
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE=99145366; PubMed=10022768;
RA   Teo B.Y., Tan N.S., Lim E.H., Lam T.J., Ding J.L.;
RT   "A novel piscine vitellogenin gene: structural and functional analyses
RT   of estrogen-inducible promoter.";
RL   Mol. Cell. Endocrinol. 146:103-120(1998).
DR   EMBL; AF017250; AAD01615.1; -.
DR   EMBL; AF072686; AAD48085.1; -.
DR   INTERPRO; IPR001747; -.
DR   INTERPRO; IPR001846; -.
DR   PFAM; PF00094; vwd; 1.
DR   PFAM; PF01347; Vitellogenin_N; 1.
KW   Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17   1788       VITELLOGENIN.
SQ   SEQUENCE   1788 AA;  195001 MW;  032A27SD45421C9A CRC64;
```

Query Match
Best Local Similarity   85.4%;   Score 82; DB 13; Length 1788;
                         95.0%;   Pred. No. 0.0031;
Matches   19; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 MRVLVLALAVALAVGDGSNL 20
          ||||||||||||||||| |||
Db   1 MRVLVLALAVALAVGDQSNL 20

RESULT   2
ID   Q9XIV8
AC   Q9XIV8;   PRELIMINARY;   PRT;   330 AA.
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PEROXIDASE (EC 1.11.1.7).
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC   Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. SAMSUN NN; TISSUE=TMV-INFECTED LEAF;
RA   Hiraga S., Ito H., Matsui H., Honma M., Ohashi Y.;
RT   "cDNA Sequences for Two Novel Tobacco Peroxidase Isoenzymes.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB027753; BAA82307.1; -.
HSSP; P00433; 2ATJ.
INTERPRO; IPR000823; -.
INTERPRO; IPR002016; -.
INTERPRO; IPR002207; -.
PFAM; PF00141; peroxidase; 1.
PRINTS; PR00458; PEROXIDASE.
PRINTS; PR00459; ASPEROXIDASE.
PRINTS; PR00461; P1PEROXIDASE.
PROSITE; PS00435; PEROXIDASE_1; 1.
PROSITE; PS00436; PEROXIDASE_2; 1.
KW   Peroxidase; Oxidoreductase.
SQ   SEQUENCE   330 AA; 35731 MW; 2034ACB6376B6180 CRC64;

Query Match
Best Local Similarity   50.0%;   Score 48; DB 10; Length 330;
                         42.9%;   Pred. No. 34;
Matches   9; Conservative   6; Mismatches   6; Indels   0; Gaps   0;

Qy   1 MRVLVLALAVALAVGDGSNLG 21
          | :||||| | :.| : :| :|
Db   15 MVILVLAIDVTMVLGQGTRVG 35

RESULT   3
ID   Q9W4U2
AC   Q9W4U2;   PRELIMINARY;   PRT;   176 AA.
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   CG3603 PROTEIN.
GN   CG3603.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-PfannKoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Andrews-PfannKoch C., Baldwin D.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari P., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003426; AAF45844.1; -.
HSSP; P08074; 1CYD.
FLYBASE; FBgn0029648; CG3603.
INTERPRO; IPR000276; -.
INTERPRO; IPR002198; -.
INTERPRO; IPR002347; -.
PFAM; PF00106; adh_short; 1.
PRINTS; PR00080; SDRFAMILY.
PRINTS; PR00081; GDHRDH.
PROSITE; PS00061; ADH_SHORT; 1.
PROSITE; PS00237; G_PROTEIN_RECEPTOR; UNKNOWN_1.
SQ   SEQUENCE   176 AA; 18567 MW; 4F0D8EAEA294324 CRC64;

Query Match
Best Local Similarity   46.9%;   Score 45; DB 5; Length 176;
                         57.1%;   Pred. No. 48;
Matches   12; Conservative   1; Mismatches   8; Indels   0; Gaps   0;

Qy   1 MRVLVLALAVALAVGDGSNLG 21
          : |.||| ||| .| ::| :|
Db   1 MSVGVLAGKVALVTGAGSGIG 21

RESULT   4
ID   Q9LV78
AC   Q9LV78;   PRELIMINARY;   PRT;   262 AA.
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   GENOMIC DNA, CHROMOSOME 5, P1 CLONE:MXK3.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.

RC   STRAIN-COLUMBIA;
RA   Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT   features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:31-63(2000).
DR   EMBL; AB019236; BAA97312.1; -.
SQ   SEQUENCE   262 AA; 28098 MW; 5CD39764789J8A11 CRC64;

Query Match   46.9%; Score 45; DB 10; Length 262;
Best Local Similarity   56.2%; Pred. No. 72;
Matches   9; Conservative   4; Mismatches   3; Indels   0; Gaps   0;

Qy   4 LVLALAVALAVGDGSN 19
            |:: || |:||:|||
Db   82 LLIMAAVLLSVGEGSN 97

RESULT 5
O18118
ID   O18118   PRELIMINARY;   PRT; 330 AA.
AC   O18118;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   T23F1.6 PROTEIN.
GN   T23F1.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditina;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wilkinson J.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan R., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z81129; CAB03405.1; -.
SQ   SEQUENCE   330 AA; 36605 MW; F043B1A90D3A8FE9 CRC64;

Query Match   46.9%; Score 45; DB 5; Length 330;
Best Local Similarity   45.0%; Pred. No. 90;
Matches   9; Conservative   5; Mismatches   6; Indels   0; Gaps   0;

Qy   1 MRVLVLALAVALAVGDGSNL 20
            || ::: |||||| |::
Db   1 MRCVILFAVALAVAQASSI 20

RESULT 6
Q9UT43
ID   Q9UT43   PRELIMINARY;   PRT; 691 AA.
AC   Q9UT43;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)

DE   PUTATIVE ATPASE (FRAGMENT).
GN   SPAC821.13C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC   Schizosaccharomycetaceae; Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Rieger M., Lyne M., Rajandream M.A., Barrell B.G.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL121770; CAB57447.1; -.
DR   INTERPRO; IPR001757; -.
DR   PFAM; PF00122; E1-E2_ATPase; 1.
PT   NON_TER   691   691
SQ   SEQUENCE   691 AA; 78275 MW; F0B8FEAA5753CC22 CRC64;

Query Match   46.9%; Score 45; DB 3; Length 691;
Best Local Similarity   64.3%; Pred. No. 1.9e-02;
Matches   9; Conservative   2; Mismatches   3; Indels   0; Gaps   0;

Qy   6 LALAVALAVGDGSN 19
            | || |:||:|:|
Db   362 LEKAVTLAIGDGAN 375

RESULT 7
Q9VGD4
ID   Q9VGD4   PRELIMINARY;   PRT; 1718 AA.
AC   Q9VGD4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG14741 PROTEIN.
GN   CG14741.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

```
Db   184 VLKNVIAVAAGMGDGLHLG 202
         ||  :||| :|||  :||
```

RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003694; AAF54749.1; -.
DR   FLYBASE; FBgn0037989; CG14741.
DR   INTERPRO; IPR000661; -.
DR   INTERPRO; IPR001757; -.
DR   PFAM; PF00122; E1-E2_ATPase; 3.
DR   PRINTS; PR00119; CATATPASE.
DR   PRINTS; PR00121; NAKATPASE.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
SQ   SEQUENCE   1718 AA; 192415 MW; EE033F1EE5CF480B CRC64;

Query Match   46.9%; Score 45; DB 5; Length 1718;
Best Local Similarity   45.8%; Pred. No. 4.7e+02;
Matches   11; Conservative   4; Mismatches   4; Indels   3; Gaps   6;

Qy   2 RVLVLAL------AVALAVGDGSN 19
          |:||:: ||||||||:||:|
Db   1270 KALVVELIKRAKNAYTLAIGDGAN 1293

RESULT 8
ID   Q9RR76   PRELIMINARY;   PRT;   328 AA.
AC   Q9RR76;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   GLYCEROL-3-PHOSPHATE DEHYDROGENASE, NAD(+)-DEPENDENT.
GN   DR2621.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome Sequence of the Radioresistant Bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE002091; AAF12158.1; -.
DR   TIGR; DR2621; -.
DR   INTERPRO; IPR001652; -.
DR   INTERPRO; IPR001993; -.
DR   INTERPRO; IPR002106; -.
DR   PFAM; PF01210; NAD_Gly3P_dh; 1.
DR   PRINTS; PR00077; GPDHDRGNASE.
DR   PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR   PROSITE; PS00139; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
SQ   SEQUENCE   328 AA; 35597 MW; 46D909D7D14170BE CRC64;

Query Match   45.8%; Score 44; DB 2; Length 328;
Best Local Similarity   52.6%; Pred. No. 1.2e+02;
Matches   10; Conservative   3; Mismatches   6; Indels   0; Gaps   0;

Qy   3 VLVLALAVALAVGDGSNLG 21

RESULT 9
ID   P96493   PRELIMINARY;   PRT;   333 AA.
AC   P96493;
DT   01-MAY-1997 (TREMBLrel. 03, Created)
DT   01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT   01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE   PUTATIVE GLYCINE DEHYDROGENASE (DECARBOXYLATING) SUBUNIT 1
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE)
DE   (GLYCINE CLEAVAGE SYSTEM P-PROTEIN) (FRAGMENT).
GN   YOHJ.
OS   Thermus aquaticus (subsp. thermophilus).
OC   Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.
OX   NCBI_TaxID=274;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HB27;
RA   Baetens M., Van de Casteele M., Legrain C., Glansdorff N.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases
CC   -! CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-
CC   AMINOMETHYLDIHYDROLIPOYLPROTEIN + CO(2).
CC   -! COFACTOR: PYRIDOXAL-PHOSPHATE.
DR   EMBL; Y11467; CAA72254.1; -.
DR   INTERPRO; IPR001023; -.
DR   PROSITE; PS01036; HSP70_3; UNKNOWN_1.
KW   Oxidoreductase.
FT   NON_TER   1   1
FT   NON_TER   333   333
SQ   SEQUENCE   333 AA; 35503 MW; 01969F09D91FE020 CRC64;

Query Match   45.8%; Score 44; DB 2; Length 333;
Best Local Similarity   60.0%; Pred. No. 1.3e+02;
Matches   9; Conservative   2; Mismatches   4; Indels   0; Gaps   0;

Qy   7 ALAVALAVGDGSNLG 21
          | :||||||:||
Db   140 AYGVDIAVGDGQSLG 154

RESULT 10
ID   Q9ROG6   PRELIMINARY;   PRT;   755 AA.
AC   Q9ROG6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   CARTILAGE OLIGOMERIC MATRIX PROTEIN PRECURSOR.
GN   COMP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CARTILAGE;
RA   Fang C., Carlson C.S., Leslie M.P., Tulli H., Stolerman E., Perris R.,
RA   Ni L., Di Cesare P.E.;
RT   "Molecular Cloning, Sequencing, Tissue and Developmental Expression of
RT   Mouse Cartilage Oligomeric Matrix Protein (COMP).";
RL   J. Orthop. Res. 0:0-0(1999).
DR   EMBL; AF033530; AAD01972.1; -.
HSSP; P35555; 1EMN
MGD; MGI:88469; Comp.
DR   INTERPRO; IPR000561; -.
DR   INTERPRO; IPR001881; -.
DR   INTERPRO; IPR002048; -.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 2.

```
KW  Signal; Matrix protein.
FT  SIGNAL        1     19          POTENTIAL.
FT  CHAIN        20    755          CARTILAGE OLIGOMERIC MATRIX PROTEIN.
SQ  SEQUENCE   755 AA;  82352 MW;  7DDFCF443589A0B7 CRC64;

Query Match             45.8%;  Score 44; DB 11; Length 755;
Best Local Similarity   66.7%;  Pred. No. 2.8e+02;
Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    3 VLVLALAVALAVGDG 17
        ||||||| | | |
Db    7 VLVLALAILRATGCG 21

RESULT 11
ID   O44385      PRELIMINARY;      PRT;   815 AA.
AC   O44385;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PUTATIVE E1-E2 ATPASE (FRAGMENT).
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CB1489 HIM-8(E1489);
RA   Halleck M.S., Pradhan D., Blackman C., Berkes C., Williamson P.L.,
RA   Schlegel R.A.,
RL   Genome Res. 0:0-0(1998);
DR   EMBL; AF034078; AAC02976.1; -.
DR   INTERPRO; IPR001757; -.
DR   PFAM; PF00122; E1-E2_ATPase; 4.
DR   PRINTS; PR00119; CATATPASE.
DR   PROSITE; PS00154; ATPASE_E1_E2; UNKNOWN_1.
FT   NON_TER       1                 .
SQ   SEQUENCE   815 AA;  91486 MW;  D5B54ED4BDFDF232 CRC64;

Query Match             45.8%;  Score 44; DB 5; Length 815;
Best Local Similarity   64.3%;  Pred. No. 3.1e+02;
Matches    9;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    6 LALAVALAVGDGSN 19
        || || ||:|||:|
Db  431 LAKHVLAIGDGAN 444

RESULT 12
ID   Q9WW45      PRELIMINARY;      PRT;   870 AA.
AC   Q9WW45;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   FIMBRIAL USHER.
GN   BCFC.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 14028;
RA   Tsolis R.M., Townsend S.M., Ficht T.A., Adams L.G., Baumler A.J.;
RT   "Attenuation in mice does not predict the ability of Salmonella
RT   typhimurium mutants to cause enteritis.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 14028;
RA   Townsend S.M., Tsolis R.M., Adams L.G., Ficht T.A., Baumler A.J.;
RT   "Distribution of bcf operon in Salmonella enterica.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF130422; AAD34372.1; -.
DR   EMBL; AF129435; AAD22098.1; -.
DR   INTERPRO; IPR000015; -.
DR   PFAM; PF00577; Usher; 1.
DR   PROSITE; PS01151; FIMBRIAL_USHER; 1.
SQ   SEQUENCE   870 AA;  94687 MW;  9E6B7120B507B871 CRC64;

Query Match             45.8%;  Score 44; DB 2; Length 870;
Best Local Similarity   72.7%;  Pred. No. 3.3e+02;
Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   11 ALAVGDGSNLG 21
        ||| | |:|||
Db  418 ALAIGSGANLG 428

RESULT 13
ID   Q51918      PRELIMINARY;      PRT;   992 AA.
AC   Q51918;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PROTEIN L PRECURSOR.
OS   Peptostreptococcus magnus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Peptostreptococcus.
OX   NCBI_TaxID=1260;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3316;
RX   MEDLINE=95078460; PubMed=7987012;
RA   Murphy J.P., Trowern A.R., Duggleby C.J.;
RT   "Nucleotide sequence of the gene for peptostreptococcal protein L.";
RL   DNA Seq. 4:259-265(1994).
DR   EMBL; L04466; AAA67503.1; -.
     HSSP; Q51911; 1GAB.
DR   INTERPRO; IPR001899; -.
DR   INTERPRO; IPR002988; -.
DR   INTERPRO; IPR003147; -.
DR   PFAM; PF00746; Gram_pos_anchor; 1.
DR   PFAM; PF01468; GA; 4.
DR   PFAM; PF02246; B1; 4.
KW   Signal.
FT   SIGNAL        1     24          POTENTIAL.
FT   CHAIN        25    992          PROTEIN L.
SQ   SEQUENCE   992 AA;  108699 MW;  9CFF571578A5DCE CRC64;

Query Match             45.8%;  Score 44; DB 2; Length 992;
Best Local Similarity   50.0%;  Pred. No. 3.7e+02;
Matches    9;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    2 RVLVLALAVALAVGDGSN 19
        ::||| |: |: ||:|
Db    6 KLLMAALAGAITVGGGAN 23

RESULT 14
ID   Q9U280      PRELIMINARY;      PRT;   1108 AA.
AC   Q9U280;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   Y49E10.11 PROTEIN.
GN   Y49E10.11.
OS   Caenorhabditis elegans.
```

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; Z98866; CAB11550.3; -.
DR INTERPRO; IPR000345; -.
DR INTERPRO; IPR001757; -.
DR PRINTS; PR00119; CATATPASE.
DR PROSITE; PS00154; ATPASE_E1_E2; UNKNOWN_1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
SQ SEQUENCE 1108 AA; 124865 MW; 298F1CD744799E2A CRC64;

Query Match   45.8%; Score 44; DB 5; Length 1108;
Best Local Similarity 64.3%; Pred. No. 4.2e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy   6 LALAVALAVGDGSN 19
            ||.|.||:|||:|
Db   755 LAKHVVLAIGDGAN 768

RESULT 15
Q9V6P5
ID Q9V6P5   PRELIMINARY;   PRT; 1297 AA.
AC Q9V6P5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG17034 PROTEIN.
GN CG17034.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003818; AAF58378.1; -.
DR HSSP; Q41261; 1QOU.
DR FLYBASE; FBgn0031837; CG17034.
DR INTERPRO; IPR001757; -.
DR PFAM; PF00122; E1-E2_ATPase; 3.
DR PRINTS; PR00119; CATATPASE.
DR PROSITE; PS00154; ATPASE_E1_E2; 1.
SQ SEQUENCE 1297 AA; 147293 MW; B11ED6E75B4D9D8A CRC64;

Query Match   45.8%; Score 44; DB 5; Length 1297;
Best Local Similarity 72.7%; Pred. No. 4.9e+02;
Matches 8; Conservative B; Mismatches 2; Indels 1; Gaps 0;

Qy   9 AVALAVGDGSN 19
            ||.||:|||:|
Db   843 AVTLAIGDGAN 853

Search completed: May 3, 2001, 15:52:43
Job time: 62 sec

```
           GenCore version 4.5
     Copyright (c) 1993 - 2000 Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:      May  5, 2001, 21:23:56 ; Search time 1035.58 Seconds
                                        (without alignments)
                                     1139.293 Million cell updates/sec

Title:       US-09-426-776-11
Perfect score: 80
Sequence:    1 attcaccatccaccagccatg........gggaccagtccaacttgggg 80

Scoring table: IDENTITY.NUC
               Gapop 10.0 , Gapext 1.0

Searched:      1288235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters:    2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
      1: gb_ba1:*                        44: em_ov:*
      2: gb_ba2:*                        45: em_pat:*
      3: gb_ba3:*                        46: em_ph:*
      4: gb_in1:*                        47: em_pl:*
      5: gb_in2:*                        48: em_ro:*
      6: gb_in3:*                        49: em_sts:*
      7: gb_om:*                         50: em_sy:*
      8: gb_ov:*                         51: em_un:*
      9: gb_pat1:*                       52: em_vi:*
     10: gb_pat2:*                       53: gb_sts1:*
     11: gb_ph:*                         54: gb_sts2:*
     12: gb_pl1:*                        55: gb_sts3:*
     13: gb_pl2:*                        56: gb_sy:*
     14: gb_pl3:*                        57: gb_un:*
     15: gb_pl4:*                        58: gb_vi1:*
     16: em_ba1:*                        59: gb_vi2:*
     17: em_ba2:*                        60: gb_htg1:*
     18: em_fun:*                        61: gb_htg2:*
     19: em_htgo_hum:*                   62: gb_htg3:*
     20: em_htgo_inv:*                   63: gb_htg4:*
     21: em_htgo_rod:*                   64: gb_htg5:*
     22: em_htg_hum1:*                   65: gb_htg6:*
     23: em_htg_hum2:*                   66: gb_htg7:*
     24: em_htg_hum3:*                   67: gb_htg8:*
     25: em_htg_hum4:*                   68: gb_htg9:*
     26: em_htg_hum5:*                   69: gb_htg10:*
     27: em_htg_hum6:*                   70: gb_htg11:*
     28: em_htg_hum7:*                   71: gb_htg12:*
     29: em_htg_hum8:*                   72: gb_htg13:*
     30: em_htg_inv1:*                   73: gb_htg14:*
     31: em_htg_inv2:*                   74: gb_htg15:*
     32: em_htg_other:*                  75: gb_htg16:*
     33: em_htg_rod:*                    76: gb_htg17:*
     34: em_hum1:*                       77: gb_htg18:*
     35: em_hum2:*                       78: gb_htg19:*
     36: em_hum3:*                       79: gb_htg20:*
     37: em_hum4:*                       80: gb_htg21:*
     38: em_hum5:*                       81: gb_htg22:*
     39: em_hum6:*                       82: gb_htg23:*
     40: em_hum7:*                       83: gb_htg24:*
     41: em_in:*                         84: gb_htg25:*
     42: em_om:*                         85: gb_pr1:*
     43: em_or:*                         86: gb_pr2:*
                                         87: gb_pr3:*
                                         88: gb_pr4:*
                                         89: gb_pr5:*
                                         90: gb_pr6:*
                                         91: gb_pr7:*
                                         92: gb_pr8:*
                                         93: gb_pr9:*
                                         94: gb_ro1:*
                                         95: gb_ro2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Match	Query Length	DB	ID	Description
	1	74	92.5	5511	8	AF017250	AF017250 Oreochrom
	2	55.8	69.8	11453	8	AF072686	AF072686.Oreochrom
	3	43.4	54.2	5166	8	FHU70826	U70826 Fundulus he
c	4	29	36.2	42805	85	AC004221	AC004221 Homo sapi
	5	28.8	36.0	945	92	HSM801971	AL137301 Homo sapi
	6	28.8	36.0	4214	89	AK023132	AK023132 Homo sapi
	7	28.8	36.0	4813	89	AK024508	AK024508 Homo sapi
	8	28.8	36.0	139505	92	HSJ591C20	AL118506 Human DNA
c	9	28.6	35.8	153713	74	AC068717	AC068717 Homo sapi
	10	28.4	35.5	10807	9	AR070469	AR070469 Sequence
	11	28.4	35.5	10807	10	I47708	I47708 Sequence 7

```
  c  12   28.4   35.5   172815   62   AC010920      AC010920 Drosophil
  c  13   28.2   35.2      840    6   S66686        S66686 n-syb=vamp
  c  14   28.2   35.2    50701   65   AC018231      AC018231 Drosophil
  c  15   28.2   35.2   146271    3   SYCSLRB       D64000 Synechocyst
  c  16   28.2   35.2   257867    4   AC005557      AC005557 Drosophil
     17   27.6   35.2   303191    4   AE003472      AE003472 Drosophil
     18   27.6   34.5      272    7   OABLG1        X07004 10.aries ge
     19   27.6   34.5      785    7   OALGLBR       X04520 Ovine mRNA
     20   27.6   34.5      824    9   AR002668      AR002668
     21   27.6   34.5      824   10   I43695        I43695 Sequence 29
     22   27.6   34.5     1769    7   SHPBLGA1      M32233 Ovis beta-1
     23   27.6   34.5     7379    7   OALGB         X12817 Ovis aries
  c  24   27.6   34.5    62838   64   AC015633      AC015633 Homo sapi
  c  25   27.4   34.2   170245   91   HS109F14      AL022721 Human DNA
  c  26   27.2   34.0   190289   59   MCU60315      U60315 Molluscum c
  c  27   27.2   34.0   211123   81   AL450163      AL450163 Homo sapi
  c  28   27     33.8   144096   80   AL359312      AL359312 Homo sapi
  c  29   27     33.8   144981   82   AL513014      AL513014 Homo sapi
  c  30   27     33.8   152044   73   AC068255      AC068255 Homo sapi
     31   27     33.8   152211    7   SSC251829     AJ251829 Sus scrof
     32   26.8   33.5      902   56   XXU49832      U49832 Synthetic s
     33   26.8   33.5     4186    9   AX026418      AX026418 Sequence
     34   26.8   33.5     4187    9   AX026417      AX026417 Sequence
     35   26.8   33.5     4187   93   HUMRSC390     D13643 Homo sapien
     36   26.8   33.5     4248    9   AX026416      AX026416 Sequence
     37   26.8   33.5     4248   88   AF261758      AF261758 Homo sapi
     38   26.8   33.5    31568    9   AC020956      AC020956 Homo sapi
  c  39   26.8   33.5   151802   79   AL354867      AL354867 Homo sapi
  c  40   26.8   33.5   163479   62   AC010616      AC010616 Homo sapi
  c  41   26.8   33.5   169072   87   AC009946      AC009946 Homo sapi
     42   26.6   33.2     2248   14   PAB302017     PAB302017 Picea abi
     43   26.6   33.2   243238   74   AC069308      AC069308 Mus muscu
  c  44   26.4   33.0     3391   92   HSG9A         X69838 H.sapiens m
  c  45   26.4   33.0     7708   59   SOUCAPPRO     L07418 Southampton
```

ALIGNMENTS

RESULT   1
AF017250
LOCUS   AF017250   5511 bp   mRNA   VRT   05-JAN-1999
DEFINITION Oreochromis aureus vitellogenin precursor (Vtg1) mRNA, complete
            cds.
ACCESSION   AF017250
VERSION   AF017250.1   GI:4102880
KEYWORDS   .
SOURCE   Oreochromis aureus
  ORGANISM Oreochromis aureus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
            Labroidei; Cichlidae; Oreochromis.
REFERENCE   1 (bases 1 to 5511)
  AUTHORS   Lim,E.H., Lam,T.J. and Ding,J.L.
  TITLE   Cloning of Full-Length Oreochromis aureus Vitellogenin cDNA and its
            Deduced Primary Structure
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 5511)
  AUTHORS   Lim,E.H., Lam,T.J. and Ding,J.L.
  TITLE   Direct Submission
  JOURNAL   Submitted (05-AUG-1997) School of Biological Sciences, National
            University of Singapore, 10 Kent Ridge Crescent 119260, Singapore
FEATURES   Location/Qualifiers
     source   1..5511
                     /organism="Oreochromis aureus"
                     /db_xref="taxon:47969"
                     /sex="female"
                     /tissue_type="liver"
                     /dev_stage="estradiol-induced adult"
                     /clone="poaVtg1"
     5'UTR   1..13
                     /gene="Vtg1"

```
     gene            1..5511
                     /gene="Vtg1"
     sig_peptide     14..61
                     /gene="Vtg1"
     CDS             14..5380
                     /gene="Vtg1"
                     /function="egg yolk protein precursor"
                     /codon_start=1
                     /product="vitellogenin precursor"
                     /protein_id="AAD01615.1"
                     /db_xref="GI:4102881"
                     /translation="MRVLVLALAVALAVGDQSNLAPGFASVKTVMYKYEAVLMGGLPE
                     EGLARAGVKIRGKVLISATSANDYILKLFSPIWPKDPFHPATKLTTALATQ
                     LSTPIKFEYTNGVVGRLAAPPGVSTTVLNIYRGIINLLQLNVKKTQNYVEMQESGAHG
                     VCKTNYVIREDARAERIHLTKTKDLNHCQEKIMKAIGLEHVERCHDCEARGKSLKGTA
                     SYNIMKPAPSGSLIMEAVAREVEFSPFNILNGAAQMESKQILTFLDIENTPVDHAR
                     YTYVHRGSLQYEHGSEILQTPIHLLRVTHAEAQIVSTLNHLVASNVAKVHEDAPLKFV
                     ELIQVMRVARFEIESLNAQFKSRPDHRYWLLNAVPHIRTHAALKFLIEKLLANELSE
                     TEAAMALLECLHSVTADQKTIELVRSLAENHRVKRNAVLNEIVMLGWGTVISRFCKAQ
                     PSCSSDLVTPVHRQVAEAVETGDIDQLVTLKCLDNAGHPASIKTIMKFLPGFGSAAA
                     RVPLKVQVDAVLALRRIAKREPRMVQEIAAQLLMEKHLHAELRMVAAMVLFETKLPVG
                     LAASISTALIKEKNLQVVSFVYSYMKAMAKTTSPDHVSVAAACNVALRFLNPKLGRLN
                     FRYSRAFHVDTYNNAWMMGAAASAVLINDAATVLPRMIMAKARTYMAGAYVDAFEVGV
                     RTEGIQEALLKRRHENSENADRITKIKQAMRALSEWRANPSSQALASMVKVFGQEIA
                     FANIDKSKVDQLIQFASGPLRVFRDAVNVSLSGYATHFAKPMLLGELRLLPTTVGL
                     PMEISLITSAVTAASVDVQATVSPPLPVNYRVSQLLESDIQLRATVAPSLAMQTYAFM
                     GVNTALIQAAVMTKAKVYTAVPAQIKARIDIVKGNLKVEFLSLQGINTIASAHAETVA
                     IARNVEDLPAARSTPLISSETASQLSKASLNSKISRMASSVTGMSASSEIIPADLPS
                     KIGRKMKLPKTYRKKIRASSRMLGFKAYAEIESHNAAYIRDCPLYALIGKHAASVRIA
                     PASGPVIEKIEVEIQVGDKAAENMIKAIDMSEEEEALEDKNVLLKIKKILAPGLKNTT
                     SSSSSSSSSSSSSSSNNKSSSSSSRSSQSSSSRSHRSRKSQSSSSQSSRSPSPSSS
                     SSSSSSSSRSSSSRSSRSSRSSRSSRSSRSSRSRSSSSSSSSSKNSKRSKRSNSKSS
                     NSSSSSSASYLLSSSKRRSRSRSSRSSRSSRSSRSARSFDSIYNKARYLANTL
                     SSRSSRRSAQSKQQLLALKFRKNHVHRHAISTORGSSHSSARSFDSIYNKARYLANTL
                     TPAMSIAIRAVRVDHKVQGYQLAAYLDKQTNRLQLIFARVAEKDNWRICADIVQLSSH
                     KLMAKTAWGAECKQYSTMIVAETGLLGHEPAARLKLTWDKLPGSIKHYAKRALKSIVP
                     IAQEYGVNYAKAKNPRNQIKLTVAVATETSMNIVLNTPKAIVYKRGVCLPVALPIGNT
                     AAELQATRDNWADKMSYLVTKANAVECSLINNTLTTFNNRKARDELPHSCYQVLAQDC
                     TPELKFMVLLKKDQIOQDQNQINVKISDIDVDMYRKNNAIAVMWNGVEIPNSNLPYLHP
                     SGNIHIRQSNEGITLNAPSHGLQEVFLGFNELRVKVADMWGKTCGACTASGNVGDE
                     YRTPSEQVTKDAISYAHSWVLSSNTCRDPSECSIKQESVKLEKRVIPEGVESKCYSVE
                     PVLQCLPGCIPVRTTVNVGFHCLPSDTTVDRSGLSSFFEKSIDLRDTAEAHLACRCT
                     PQCA"
     mat_peptide     62..5377
                     /gene="Vtg1"
                     /product="vitellogenin"
     misc_feature    3269..3946
                     /gene="Vtg1"
                     /note="encodes putative phosvitin domain"
     3'UTR           5381..5511
                     /gene="Vtg1"
     polyA_signal    5465..5470
                     /gene="Vtg1"
```

BASE COUNT   1474 a   1443 c   1299 g   1295 t
ORIGIN

Query Match   92.5%;   Score 74; DB 8; Length 5511;
Best Local Similarity   100.0%; Pred. No. 7.7e-15;
Matches   74; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   5 acatccaccagccatgagggtgcttgtactagctcctggctccgcagtggggga 64
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ACATCCACCAGCCATGAGGGTGCTTGTACTAGCTCCTGGCTCCGCAGTGGGGGA 60

Qy   65 ccagtccaacttgg 78
          ||||||||||||||
Db   61 CCAGTCCAACTTGG 74

RESULT   2
AF072686
LOCUS   AF072686   11453 bp   DNA   VRT   29-NOV-1999

DEFINITION Oreochromis aureus vitellogenin gene, complete cds.
ACCESSION   AF072686 X93569 X93570 X93571 271336
VERSION   AF072686.1 GI:5725515
KEYWORDS   .
SOURCE   Oreochromis aureus
ORGANISM   Oreochromis aureus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
            Labroidei; Cichlidae; Oreochromis.
REFERENCE   1 (bases 1 to 11453)
AUTHORS   Teo,B.Y., Tan,N.S., Lim,E.H., Lam,T.J. and Ding,J.L.
TITLE   A novel piscine vitellogenin gene: structural and functional
            analyses of estrogen-inducible promoter
JOURNAL   Mol. Cell. Endocrinol. 146 (1-2), 103-120 (1998)
MEDLINE   99145366
PUBMED   10022768
REFERENCE   2 (bases 1 to 11453)
AUTHORS   Teo,B.Y., Lim,E.H. and Ding,J.L.
TITLE   Direct Submission
JOURNAL   Submitted (18-JUN-1998) Department of Biological Sciences, National
            University of Singapore, 10 Kent Ridge Crescent, Singapore 119260,
            Singapore
COMMENT   On Nov 29, 1999 this sequence version replaced gi:1771394
            gi:1771393 gi:1771392 gi:1403509.
FEATURES   Location/Qualifiers
     source   1..11453
                     /organism="Oreochromis aureus"
                     /db_xref="taxon:47969"
                     /clone="Oavg1"
                     /sex="female"
                     /tissue_type="blood"
                     /dev_stage="adult"
     promoter   1..1677
     protein_bind   141..146
                     /note="GATA binding site"
                     /bound_moiety="GATA factor"
     protein_bind   315..320
                     /note="GATA binding site"
                     /bound_moiety="GATA factor"
     protein_bind   326..338
                     /note="GATA binding site"
                     /bound_moiety="GATA factor"
     protein_bind   353..358
                     /note="estrogen response element (ERE); imperfect ERE"
                     /bound_moiety="estrogen receptor protein"
     protein_bind   440..445
                     /note="GATA binding site"
                     /bound_moiety="GATA factor"
     protein_bind   1001..1006
                     /note="GATA binding site"
                     /bound_moiety="GATA factor"
     protein_bind   1146..1158
                     /note="estrogen response element (ERE); imperfect ERE"
                     /bound_moiety="estrogen receptor protein"
     protein_bind   1200..1205
                     /note="GATA binding site"
                     /bound_moiety="GATA factor"
     TATA_signal   1649..1655
     5'UTR   1678..1692
     mRNA   join(1678..1732,2065..2085,2193..2344,2441..2692,
                     2793..2954,3044..3192,3280..3430,3519..3676,3755..3920,
                     4000..4125,4211..4429,4538..4755,4850..4947,5035..5159,
                     5288..5392,5502..5621,5709..5892,5990..6083,6177..6362,
                     6462..6612,6694..6910,7017..7744,7834..7878,7957..8025,
                     8117..8323,8407..8549,8736..8874,8978..9107,9224..9407,
                     9497..9591,9686..9780,9906..10071,10213..10374,
                     10471..10690)
                     /product="vitellogenin"
     CDS   join(1693..1732,2065..2085,2193..2344,2441..2692,
                     2793..2954,3044..3192,3280..3430,3519..3676,3755..3920,
                     4000..4125,4211..4429,4538..4755,4850..4947,5035..5159,
                     5288..5392,5502..5621,5709..5892,5990..6083,6177..6362,

6462..6612,6694..6910,7017..7744,7834..7878,7957..8025,
8117..8323,8407..8549,8736..8874,8978..9107,9224..9407,
9497..9591,9686..9780,9906..10071,10213..10374,
10471..10690)
                     /note="phospholipoglycoprotein; yolk protein precursor"
                     /codon_start=1
                     /product="vitellogenin"
                     /protein_id="AAD48085.1"
                     /db_xref="GI:5725516"
                     /translation="MRVLVLALAVALAVGDQSNLAPGFASVKTYMYKYEAVLMGGLPE
                     EGLARAGVKIRGKVLISATSANDYILKLVDPQLLEYSGIWPKDPFHPATKLTTALATQ
                     LSTPIKFEYTNGVGRLAAPPGVSTTVLNIYRGIINLLQLNVKKTQNVYEMQESGAHG
                     VCKTNTVIREDARAERIHLTKTKDLNHCQEKIMKAIGLEHVEKCHDCEARGKSLKGTA
                     SYNYIMKPAPSGSLIMEAVAREVIEFSPFNLNGAQMESKQILTFLDIENTPVDHAR
                     YTYVHRGSLQYEHGSEILQTPIHLLRVTHARAQIVSTLNHLVASNVAKVHEDAPLKFV
                     ELIQVMRVARFETIESLWAQFKSRPDHRYWLLNAVPHIRTHAALKFLIEKLLANELSE
                     TEAAMALLECLHSVTADQKTIELVRSLAENHRVKRNAVLNEIVMLGWGTVINRFCKAQ
                     PSCSSDLVTPVHRQVAEAVETGDIDQLTVTLKCLDNAGHPASIKTIMKFLPGFGSAAA
                     RVPLKVQVDAVLALRRIAKREPKMVQETAAQLLMEKHLHAELRMVAAMVLFETKLPVG
                     LAASISTALIKEKNLQVVSPVYSYMKAMAKTSPDHVSVAAACNVALRFLNPKLGRLN
                     FRYSRAFHVDTYNNAWMGAAASAVLINDAATVLPRMINAKARTYMAGAYVDAFEVGV
                     RTEGIQEALLKRRHENSENADRITKIKQAMRALSEWRANPSSQALASMVKVFQEIA
                     FANIDRSKVDQLIQFASGPLRNVFRDAVNSVLSGYATHFAKPMLLGELRLILPTTVGL
                     PMEISLITSAVTAASVDVQATVSPPLPVNYRVSQLLESDIQLRATVAPSLAMQTFAFM
                     GVNTALIQAAVMTKAKVYTAVPAQIKARIDIVKGNKLKVEFSLSLQGINTIASAHAETVA
                     IARNVEDLPAARSTPLISSETASQLSKASLNSKISRMASSVTGGMSASSEIIPADLPS
                     KIGRKMKLPKTYRKKIRASSRMLGFKAYAEIESHNAAYIRDCPLYALIGKHAASVRIA
                     PASGPVIEKIEVEIQVGDKAAENMIKAIDMSEEEEALEDKNVLKIKKLLAPGLKNTT
                     SSSSSSSSSSSSSSSNKSSSSSRSSSSSSSSRHRSRSRKSQSSSSQSSQSSSRSPSSS
                     SSSSSSSRSSRSSSSSRSSSRSSSSSRSSRRTKMADIVAPIITTSRVSSSSRSRSAS
                     NSSSSSASYLLSSSKRRSRSRSSSSSSSSSSSSSSSSSSKNSKRSKSSNKSS
                     SSRSSRRSAQSKQQLLALKPRKNVHRHAISTQRGSHSSARSFDSIYNAKYLANTL
                     TPAMSIAIRAVRVDHKVQGYQLAAYLDKQTNRLQLIFARVAEKDNWRICADIVQLSSH
                     KLMAKTAWGAECKQYSTMIVAETGLLGHEPAARLKLTWDKLPGSIKHYAKRALKSIVP
                     IAQEYGVNYAKAKNPRNQIKLTVAVATETSMNIVLNTPKAIVYKRGVCLPVALPIGNT
                     AAELQATRDNWADKMSYLVTKANAVECSLINNTLTTFNNRKARDELPHSCYQVLAQDC
                     TPELKFMVLLKKDQIQDQNQINVKISDIDVDMYRKNNAIAVMVNGVEIPNSNLPYLHP
                     SGNIHIRQSNEGITLNAPSHGLQEVFLGFNELRVKVADWMKGTKCGACGTASGNVGDE
                     YRTPSEQVTKDAISYAHSWVLSSNTCRDFSECSIKQESVKLEKRVIFEGVESKCYSVE
                     PVLQCLPGCIPVRTTVNVGFHCLPSDTTVDRSGLSSFFEKSIDLRDTAEAHLACRCT
                     PQCA"
     protein_bind   2072..2084
                     /note="estrogen response element (ERE); imperfect ERE"
                     /bound_moiety="estrogen receptor protein"
     3'UTR   10581..10690
     polyA_signal   10665..10670
     polyA_site   10690
     polyA_signal   10739..10744
BASE COUNT   3340 a   2463 c   2285 g   3365 t
ORIGIN

Query Match   69.8%; Score 55.8; DB 8; Length 11453;
Best Local Similarity 89.6%; Pred. No. 1.2e-08;
Matches 60; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy   3 tcacatccaccagccatgagggtgcttgtactagctcttgctggctccgcagtgggg 62
            ||||||||||||||||||||||||||||||||||||||||||| | | |||||| ||
Db   1678 TCACATCCACCAGCCATGAGGGTGCTTGTACTAGCTCTTGCTGGCTCCGCCAGGTGAG 1737

Qy   63 gaccagt 69
            | | |
Db   1738 TAGGAAT 1744

RESULT 3
FHU70826
LOCUS   FHU70826   5166 bp   mRNA   VRT   18-OCT-1996
DEFINITION Fundulus heteroclitus vitellogenin II precursor mRNA, complete cds.
ACCESSION   U70826
VERSION   U70826.1 GI:1621358
KEYWORDS   .
SOURCE   killifish.

```
ORGANISM   Fundulus heteroclitus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
           Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
           Cyprinodontiformes; Fundulidae; Fundulus.
REFERENCE  1  (bases 1 to 5166)
  AUTHORS  LaFleur,G.J. Jr., Byrne,B.M., Haux,C., Greenberg,R.M. and
           Wallace,R.A.
  TITLE    Liver-derived cDNAs: Vitellogenins and vitelline envelope protein
           precursors (choriogenins)
  JOURNAL  Int. Symp. Reprod. Physiol. Fish 5, 336-338 (1995)
REFERENCE  2  (bases 1 to 5166)
  AUTHORS  LaFleur,G.J. Jr., Hoch,K.L., Denslow,N., Byrne,B.M. and
           Wallace,R.A.
  TITLE    Derivation of oocyte and egg proteins from parental vitellogenins
           in Fundulus heteroclitus: sequence alignments reveal alternative
           processing and a PEST site in the largest yolk protein
  JOURNAL  Unpublished
REFERENCE  3  (bases 1 to 5166)
  AUTHORS  LaFleur,G.J. Jr., Hoch,K.L., Denslow,N., Byrne,B.M. and
           Wallace,R.A.
  TITLE    Direct Submission
  JOURNAL  Submitted (15-SEP-1996) Whitney Laboratory and Dept of Anatomy and
           Cell Biology, University of Florida, 9505 Ocean Shore Blvd, St.
           Augustine, FL 32086-8623, USA
FEATURES             Location/Qualifiers
     source          1..5166
                     /organism="Fundulus heteroclitus"
                     /db_xref="taxon:8078"
                     /tissue_type="estrogen-induced liver"
     CDS             13..5076
                     /note="Vtg II; yolk protein precursor induced by estrogen;
                     mRNA levels 10-50 times less abundant than Vtg I mRNA"
                     /codon_start=1
                     /product="vitellogenin II precursor"
                     /protein_id="AAB17152.1"
                     /db_xref="GI:1621359"
                     /translation="MRVLVLALTVALVAGNQVSYAPEFAPGKTYEYKYEGYILGGLPE
                     EGLAKAGVKIQSKVLIGAAGPDSYILKLEDPVISGYSGIWPKEVFHPATKLTSALSAQ
                     LLTPVKFEYANGVIGKVFAPPGISTNVLNVFRGLLNMFQMNIKTQNVDLQETGVKG
                     VCKTHYILHEDSKADRLHLTKTTDLNHCTDSIHMDVGMAGYTEKCAECMARGKTLSGA
                     ISVNYIMKPSASGTLILEATATELLQYSPVNIVNGAVQMEAKQTVTFVDIRKTPLEPL
                     KADYIPRGSLKYELGTEFLQTPIQLLRITNVEAQTVESLNNLVSLNMGHAHEDSPLKF
                     IELIQLLRVAKYESIEALWSQFKTKIDHRHWLLSSIPAIGTHVALKFIKEKIVAGEVT
                     AEEAAQAIMSSTHLVKADLEAIKLQEGLAVTPNIRENAGLRELVMLGFGIMVHKYCVE
                     NPSCPSELVRPVHDIIAKALEKRDNDELSLALKVLGNAGHPSSLKPIMKLLPGPGSSA
                     SELELRVHIDATLALRKIGKREPKMIQDVALQLFMDRTLDPELRMVAVVVLFDTKLPM
                     SYRYSRAFHYDHYHNPWMLGAAASAFYINDAATVLPKNIMAKARVYLSGVSVDVLEFG
                     ARAEGVQEALLKARDVPESADRLTKMKQALKALTEWRANPSRQPLGSLIVKVLGQDVA
                     FANIDKEMVEKIIEFATGPEIRTRGKKALDALLSGYSMKYSKPMSAIEVRHIFPTSLG
                     LPMELSLYTAAVTRAASVEVQATISPPLPEDFHPAHLLKSDISMKRASVTPSVSLHTYGV
                     MGVNSPFIQASVLSRAKDHAALPKKMEARLDIVKGYFSYQFLPVEGVKTIASARLETV
                     AIARDVEGLAAAKVTPVVPYEPIVSKNATLNLSQMSYILNDSISASSELLPFSLQRQT
                     GKNKIPKIVKKMCATTYTYGIEGCVDIWSRNATFLRNFPIYAIIGNHSLLVNVTPAA
                     GPSIERIEIEVQFGEQOAAEKILKEVYLNEEEVLEDKNVLMKLKKILSPGLKNSTKAS
                     SSSGSSRSRSRSRSSSSSSSSSSSSSSSSSSSSSSSLRRNSKMLDLADPLNITSK
                     RSSSSSSSSSSSSSSSSSSKTKWQLHERNFTKDHIHQHSVSKERLNSKSSASSF
                     ESIYNKITTLSNIVSPVVTVLVRAIRADHKNQGYOIAVYYDKLTTRVQIVANLTEDD
                     NWRICSDSMMLSHHKVMTRVTWGICGCKQYNTTIVAETGRVEKEPAVRVKLAWARLPTY
                     IRDYARRVSRYISRVAEDNGVNRTKVASKPKEIKLTVAVANETSLNVTLNPTKNTFFK
                     LGWVLPFYLPINNTAAEDLAVFQGRWMDQVTYMLTKSAAAECTVVDETVTFNNRKYKT
                     ETPHSCHQVLAQDCTSEIKFIVLLKRDQTAERNEISIKIENIDVDMYPKDNAVVVKVN
                     GVEIPLTNLPYQHPTGNIQIRQREEGISLHAPSHGLQEVFLSLNKVQVKVVDVMRGQT
                     ISLEGEESKCYSVEPVWRCLPGCAPVRFTSVTVGLPCVSLDSNLNRSDSLSSIYQKSV
                     DVSETAESHLACRCTPQCA"
     misc_feature    58..117
                     /note="encodes region of N-terminal amino acids of a 69
                     kDa-yolk protein (YP 69) that were experimentally
                     determined by Edman degradation"
     mat_peptide     58..5073
                     /product="vitellogenin II"
     misc_feature    3235..3660
                     /note="encodes Vtg polyserine domain found in
                     phosphoserine rich yolk proteins termed phosvitin and/or
                     phosvettes"
     polyA_signal    5135..5140
BASE COUNT     1360 a     1228 g     1358 c     1220 t
ORIGIN
```

```
Query Match                54.2%;  Score 43.4;  DB 8;  Length 5166;
  Best Local Similarity     76.8%;  Pred. No. 0.0019;
  Matches  53;  Conservative  0;  Mismatches  16;  Indels  0;  Gaps  0;

Qy          7 atccacagccatgagggtgctgactagctcttgtggctctcgcagtggggacc 66
              || ||||||||||||||||| |||||  ||||| ||||||| || | ||| |||
Db          2 ATTCACCAGCCATGAGGGTGCTTGTGCTTCCACTGTGGCCCTTGTGGCCGGGAACC 61

Qy         67 agtccaact 75
              || | ||
Db         62 AGGTGAGCT 70
```

```
RESULT 4
AC004221/c
LOCUS       AC004221              42805 bp    DNA                    PRI       03-AUG-2000
DEFINITION  Homo sapiens DNA from chromosome 19, cosmid R29144 (LLNLR-252D12)
            and overlapping PCR product, complete sequence.
ACCESSION   AC004221
VERSION     AC004221.2  GI:9665053
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 42805)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Adanson,A.W.,
            Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,A., Garnes,J., Danganan,L., Poundstone,P.,
            Christensen,M., Georgescu,A., Avila,J., Liu,S., Attix,C.,
            Andreise,T., Trankheim,M., Amico-Keller,G., Coefield,J., Duarte,S.,
            Lucas,S., Bruce,R., Thomas,P., Quan,G., Kronmiller,B., Arellano,A.,
            Brower,A., Montgomery,M., Ow,D., Nolan,M., Trong,S., Kobayashi,A.,
            Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
            serine protease gene cluster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 42805)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
REFERENCE   3  (bases 1 to 42805)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2000) DOE Joint Genome Institute, Lawrence
            Livermore National Laboratory, Livermore, CA 94550
COMMENT     On Aug 3, 2000 this sequence version replaced gi:2911257.
            Map and sequence oriented from p-telomere to centromere. Bases
            1-1791 from PCR product; bases 1385-42805 from clone R29144. This
            sequence overlaps cosmid R30295 (LLNLR-264D1, AC011544) on the left
            from bases 1 to 281 of this accession and overlaps cosmid R33114
            (LLNLR-293G10, AC004258) on the right from bases 40533 to 42805 of
            this accession. Additional chr19 map and sequence information are
            available at: http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES             Location/Qualifiers
     source          1..42805
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19p13.3 between CDC34 and D19S342"
                     /clone="R29144"
                     /cell_line="5HL2-B"
                     /clone_lib="LL19NC03 R chromosome 19-specific cosmid
```

```
                library"
                /note="Cosmid library constructed at LLNL from flow-sorted
                chromosomes from hybrid 5HL2-B, which carries chromosome
                19 as its only human chromosome."
gene            <1..2865
                /gene="STK11"
                /note="serine/threonine kinase 11 (Peutz-Jeghers
                syndrome); LKB1; PJS; Serine/threonine protein kinase 11"
mRNA            join(<1225..1434,2364..2865)
                /gene="STK11"
CDS             <1225..1418
                /gene="STK11"
                /note="Serine/threonine kinase 11"
                /codon_start=3
                /product="STK11 [amino acids 371-433]"
                /protein_id="AAF97257.1"
                /db_xref="GI:9665054"
                /translation="QVPEEEASHNGQRRGLPKAVCMNGTEAAQLSTKSRAEGRAPNPA
                RKACSASSKIRRLSACKQQ"
misc_feature    2356..2791
                /note="DDS similarity to overlapping ESTs:
                (972..1407) AA436398 zv44a08.s1 Soares ovary tumor NbHOT
                Homo sapiens cDNA clone 756470 3' similar to SW:VNUA_PRVKA
                P33485 PROBABLE NUCLEAR ANTIGEN. ; (433..1); 91%
                identity.-(1359..1050) AA576906 nm73e05.s1 NCI_CGAP_Co9
                Homo sapiens cDNA clone IMAGE:1073888. Score: 513
                Identity: 291/305 (95%).-(1407..1098) AA292429 zt28g10.s1
                Soares ovary tumor NbHOT Homo sapiens cDNA clone 714498
                3'. Score: 491 Identity: 290/299 (96%).-(1379..1107)
                AA642427 ns27d07.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone
                IMAGE:1184845. Score: 482 Identity: 265/269
                (98%).-(1406..1144) AA557240 n175d11.s1 NCI_CGAP_Br2 Homo
                sapiens cDNA clone IMAGE:1056501. Score: 447 Identity:
                250/260 (96%)."
repeat_region   2683..2714
                /rpt_family="GC_rich"
repeat_region   complement(2792..2841)
                /rpt_family="POLY_A"
misc_feature    2934..3240
                /note="DDS similarity to overlapping ESTs:
                (1550..1799) AA652409 ns57d10.s1 NCI_CGAP_Pr22 Homo
                sapiens cDNA clone IMAGE:1187731. Score: 478 Identity:
                247/248 (99%).-(1856..1674) W52572 zc45h09.s1 Soares
                senescent fibroblasts NbHSF Homo sapiens cDNA clone 325313
                3' (192..11); 99% identity.-(1675..1793) AA568599
                nm21c12.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone
                IMAGE:1060822. Score: 234 Identity: 118/119 (99%)."
repeat_region   complement(3293..3334)
                /rpt_family="GC_rich"
repeat_region   3482..3536
                /rpt_family="GC_rich"
repeat_region   3775..3875
                /rpt_family="SUBTEL_sat"
misc_feature    complement(3950..4913)
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: good, score: 73.000"
CDS             complement(join(3950..4913,5872..6046,8197..8407,
                8962..9102,9342..9598,9772..9916,10272..10336,
                10550..10689,10767..10853))
                /codon_start=1
                /product="Hypothetical 75.8 kDa human protein"
                /protein_id="AAC04305.1"
                /db_xref="GI:2911259"
                /translation="MATAATTTTTTATVALTTSWDNATGRPTAEPDPILDNYVLLVV
                VMSLFVGGTLVVLSGVLLCKRCWDVHQRLNRAMEEAEKTTTYLDNGTHPAQDPDFR
                GEDPECQDAETERFLST'SSTGRRVSFNEAALFEQSRKTQDKGRQGWQSTAAGASGMG
                GGRALMARCCRRYTLFEGDFHHLKNARITHLHLPPLKIVTIHECDSGEASSATTPHPA
                TSPKATLAIFQPPGKALTGRSVGPSSALPGDPYNSAAGATDFAEISPSASSDSGEGTS
                LDAGTRSTKAGGPGAAAGPGEAGPGSGAGTVLQFLTRLRRHASLDGASPYFKVKKWKL
                EPSQRAASLDTRGSPKRHHFQRQRAASESTEQEEGDAPQEDFIQYIARAGDAVAFPHP
                RPFLASPPPALGRLEAEAAGGASPDSPPERGAGSAGPEQQQPPLEPDAERDAGPEQA

                QTSYRDLWSLRASLEHAAASDHSSSGNDROSVRSGDSSGSGSGAAPAFPPPSPPAP
                RPKDGEARRLLQMDSGYASIEGRGAGDDTEPPAAPARPSRAWPRRPRRDYSIDEKT
                DALFHEFLRHDPHFDDTPAAARHRARAHPHARKQWQRGRQHSDPGARARAPALAGTPAP
                PAGAARPARAPLRRGDSVDGPPDGRTLGGAGDDPAIPVIEEPGGGCPGSGLCVLPS
                GSVLDKLAAGLDERLFPPRLAEPVVATPALVAAAPTSPDHSPA"
repeat_region   4099..4301
                /rpt_family="GC_rich"
repeat_region   4454..4513
                /rpt_family="(CGGG)n"
repeat_region   complement(4576..4650)
                /rpt_family="GC_rich"
misc_feature    complement(5872..6046)
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 2, quality: excellent, score: 80.000"
misc_feature    complement(5968..6046)
                /note="DDS similarity to overlapping ESTs:
                (4584..4662) W52572 zc45h09.s1 Soares senescent
                fibroblasts NbHSF Homo sapiens cDNA clone 325313 3';
                (271..193); 96% identity.-(4662..4533) W52571 zc45h09.r1
                Soares senescent fibroblasts NbHSF Homo sapiens cDNA clone
                325313 5' (253..393); 67% identity.
                "
repeat_region   complement(6557..6585)
                /rpt_family="MIR"
repeat_region   complement(7506..7663)
                /rpt_family="MER58B"
misc_feature    complement(8197..8407)
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: good, score: 74.000-DDS similarity to
                overlapping ESTs:
                (7023..6813) W52571 zc45h09.r1 Soares senescent
                fibroblasts NbHSF Homo sapiens cDNA clone 325313 5'
                (45..252); 90% identity.-(6991..6813) W52572 zc45h09.s1
                Soares senescent fibroblasts NbHSF Homo sapiens cDNA clone
                325313 3'; (457..272); 90% identity."
misc_feature    complement(8962..9102)
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 1, quality: excellent, score: 87.000-(7621..7578)
                DDS similarity to W52571 zc45h09.r1 Soares senescent
                fibroblasts NbHSF Homo sapiens cDNA clone 325313 5'
                (1..44); 98% identity."
misc_feature    complement(9342..9598)
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 2, quality: excellent, score: 77.000"
misc_feature    complement(9772..9916)
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 2, quality: excellent, score: 100.000"
repeat_region   complement(9941..10050)
                /rpt_family="GC_rich"
repeat_region   10172..10260
                /rpt_family="(CAGA)n"
misc_feature    complement(10272..10336)
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 1, quality: excellent, score: 98.000"
misc_feature    complement(10550..10685)
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 1, quality: excellent, score: 100.000"
repeat_region   complement(10606..10659)
                /rpt_family="(TGG)n"
misc_feature    complement(10767..10853)
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 2, quality: excellent, score: 88.000"
repeat_region   10813..10864
                /rpt_family="(TGG)n"
repeat_region   11545..11656
                /rpt_family="(CGG)n"
repeat_region   complement(12493..12523)
                /rpt_family="(CGG)n"
repeat_region   12553..12627
                /rpt_family="(CGG)n"
repeat_region   12646..12913
                /rpt_family="GC_rich"
repeat_region   complement(13948..14249)
```

```
        repeat_region    /rpt_family="AluJb"
                         complement(15011..15068)
        repeat_region    /rpt_family="GC_rich"
                         complement(15443..15762)
                         /rpt_family="AluSx"

Query Match            36.2%;    Score 29; DB 85;    Length 42805;
Best Local Similarity   71.7%;    Pred. No. 15;
Matches  38;  Conservative   0; Mismatches   15; Indels    0;  Gaps    0;

Qy      23 ggtgcttgtactagctcttgtctggctgcctcgcagtggggaccagtccaact 75
           |||||| ||||| | || || | || || |||||||| ||||||| || |||||
Db   29726 GGTGCTTGGCACCAGACTGCACTGCTGCCCTGGCCCTGGGGGAGCAGTCCATCT 29674


RESULT 5
HSM801971
LOCUS         HSM801971    945 bp    mRNA           PRI    18-FEB-2000
DEFINITION    Homo sapiens mRNA; cDNA DKFZp434N1429 (from clone DKFZp434N1429);
              partial cds.
ACCESSION     AL137301
VERSION       AL137301.1  GI:6807764
KEYWORDS      .
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 945)
AUTHORS       Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE         Direct Submission
JOURNAL       Submitted (15-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152
              Martinsried, GERMANY
COMMENT       Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
              Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
              sequenced by BMFZ (Biomedical Research Center at the Charite,
              Berlin/Germany) within the cDNA sequencing consortium of the German
              Genome Project.
              This clone (DKFZp434N1429) is available at the RZPD in Berlin.
              Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
              Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
              information about the clone and the sequencing project is available
              at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1..945
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434N1429"
                     /clone_lib="434 (synonym: htes3). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
                     /tissue_type="testis"
     gene            1..408
                     /gene="DKFZp434N1429"
     CDS             <1..408
                     /gene="DKFZp434N1429"
                     /note="unknown"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB70686.1"
                     /db_xref="GI:6807765"
                     /translation="LWNWQKLHFCYPYQQQFASSHAVSVGCGVSVCVHMCICIYHAAG
                     VSFMLLSPGEVTSTIKGGSEMASGFLGVGQGPAGGFPSRCTEGSAEPTVCRGCVPRAL
                     PRPPPALASALQGGVPIIGAVCHAHLTQVCVPV"
        polyA_signal  890..895
        polyA_site    907
BASE COUNT    223 a    216 c    282 g    224 t
ORIGIN

Query Match            36.0%;    Score 28.8; DB 92;    Length 945;
Best Local Similarity   60.0%;    Pred. No. 18;
Matches  48;  Conservative   0; Mismatches   32; Indels    0;  Gaps    0;
```

```
Qy       1 attcacatccacagccatgagggtgctgtactagctcttgtctggctctcgcagtgg 60
           | ||| || | ||| || || ||||||| | |||||| | || || ||||||||||||
Db     702 AATCAAATGCACACGCATGGAAGCTGCTTGTCCATGCACTGGTTTGTCACTGGCTCCT 761

Qy      61 gggaccagtccaacttgggg 80
           | || || ||| |||||||||
Db     762 GGAATCCGACCGTGTTGGGG 781


RESULT 6
AK023132
LOCUS         AK023132    4214 bp    mRNA            PRI    29-SEP-2000
DEFINITION    Homo sapiens cDNA FLJ13070 fis, clone NT2RP3001777.
ACCESSION     AK023132
VERSION       AK023132.1  GI:10434913
KEYWORDS      oligo capping; fis (full insert sequence).
SOURCE        Homo sapiens teratocarcinoma cell line:NT2 cDNA to mRNA,
              clone_lib:NT2RP3 clone:NT2RP3001777.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (sites)
AUTHORS       Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
              Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
              Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
              Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
              Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
              Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Ninomiya,K. and Iwayanagi,T.
              Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
TITLE         NEDO human cDNA sequencing project
JOURNAL       Unpublished (2000)
REFERENCE     2  (bases 1 to 4214)
AUTHORS       Isogai,T. and Otsuki,T.
TITLE         Direct Submission
JOURNAL       Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
              Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
              Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
              Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT       NEDO human cDNA sequencing project supported by Ministry of
              International Trade and Industry of Japan; cDNA full insert
              sequencing: Research Association for Biotechnology; cDNA library
              construction, 5'- & 3'-end one pass sequencing and clone selection:
              Helix Research Institute (supported by Japan Key Technology Center
              etc.) and Department of Virology, Institute of Medical Science,
              University of Tokyo.
FEATURES             Location/Qualifiers
     source          1..4214
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone="NT2RP3001777"
                     /clone_lib="NT2RP3"
                     /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
BASE COUNT    779 a   1120 c   1278 g   1037 t
ORIGIN

Query Match            36.0%;    Score 28.8; DB 89;    Length 4214;
Best Local Similarity   60.0%;    Pred. No. 18;
Matches  48;  Conservative   0; Mismatches   32; Indels    0;  Gaps    0;
```

```
Qy       1 attcacatccacagccatgagggtgctgtactagctcttgtctggctctcgcagtgg 60
           | ||| || | ||| || || ||||||| | |||||| | || || ||||||||||||
Db    4009 AATCAAATGCACACGCATGGAAGCTGCTTGTCCATGCACTGGTTTGTCACTGGCTCCT 4068

Qy      61 gggaccagtccaacttgggg 80
           | || || ||| |||||||||
Db    4069 GGAATCCGACCGTGTTGGGG 4088
```

```
RESULT    7
AK024508
LOCUS        AK024508    4813 bp    mRNA                    PRI    29-SEP-2000
DEFINITION   Homo sapiens mRNA for FLJ00118 protein, partial cds.
ACCESSION    AK024508
VERSION      AK024508.1  GI:10440535
KEYWORDS     fis (full insert sequence).
SOURCE       Homo sapiens adult spleen cDNA to mRNA, clone:as00118.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 4813)
AUTHORS      Ohara,O., Nagase,T., Kikuno,R. and Okumura,K.
TITLE        The nucleotide sequence of a long cDNA clone isolated from human
             spleen
JOURNAL      Published Only in Database (2000) In press
REFERENCE    2  (bases 1 to 4813)
AUTHORS      Ohara,O., Nagase,T., Kikuno,R. and Okumura,K.
TITLE        Direct Submission
JOURNAL      Submitted (24-AUG-2000) to the DDBJ/EMBL/GenBank databases. Osamu
             Ohara, Kazusa DNA Research Institute, Department of Human Gene
             Research; 153-3, Yana, Kisarazu, Chiba 292-0812, Japan
             (E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp/NEDO
             Tel:81-438-52-3913, Fax:81-438-52-3914)
COMMENT      NEDO human cDNA sequencing project supported by Ministry of
             International Trade and Industry of Japan; cDNA full insert
             sequencing: 5'- & 3'-end one pass sequencing and clone selection:
             construction; 5'- & 3'-end one pass sequencing and clone selection:
             Kazusa DNA Research Institute.
FEATURES             Location/Qualifiers
     source          1..4813
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="as00118"
                     /dev_stage="adult"
                     /tissue_type="spleen"
                     /note="vector:pBluescriptII SK plus"
     gene            1..373
                     /gene="FLJ00118"
     CDS             <1..373
                     /gene="FLJ00118"
                     /note="Start codon is not identified."
                     /codon_start=2
                     /product="FLJ00118 protein"
                     /protein_id="BAB15798.1"
                     /db_xref="GI:10440536"
                     /translation="YDKYGSLGLVVAEQFGEENVNTYFVLSSWWAKALFVFCGLLTCC
                     YCCCLCCCFNCCCGKCKPKAPEGEETEFYVSPEDLEAQLQSDEREATDTPIVIQPAS
                     ATETTQLTADSHPSHTDGFN"
BASE COUNT    904 a   1308 c   1445 g   1156 t
ORIGIN

Query Match             36.0%;  Score 28.8;  DB 89;  Length 4813;
Best Local Similarity   60.0%;  Pred. No. 18;
Matches   48;  Conservative   0;  Mismatches   32;  Indels   0;  Gaps   0;

Qy      1 attcacatccacagccatgagggtgcttgtactagctcttgctgtggctctcgcagtgg 60
          |  ||| || || || ||  |||||||| |||||| |||||| || || ||| || ||||
Db   4603 AATCAAATGCACACGGCATGGAAGCTGCTTGTCCATGCACTGCACTGGTTTTGTCACTGGCGTCCT 4662

Qy     61 gggaccagtccaacttgggg 80
          | ||| ||  |||||| |||||
Db   4663 GGAATCCGACCGTGTTGGGG 4682


RESULT    8
HSJ591C20
LOCUS        HSJ591C20  139505 bp    DNA                    PRI    05-FEB-2001
DEFINITION   Human DNA sequence from clone RP4-591C20 on chromosome 20 Contains
             ESTs, STSs, GSSs and CpG islands. Contains a novel gene for a
             protein similar to mouse NG26, the TPD52L2 gene for three isoforms
             of tumor protein D52-like protein 2, a gene for a novel DnaJ domain
```

```
protein similar to mouse and bovine cysteine string protein with
two isoforms, a gene for a novel phosphoribulokinase with three
isoforms, the KIAA1196 gene, the 5' part of the TOM gene for a
putative mitochondrial outer membrane protein import receptor
similar to yeast pre-mRNA splicing factors Prp1/Zer1 and Prp6 and a
novel gene, complete sequence.
ACCESSION    AL118506
VERSION      AL118506.27  GI:10190557
KEYWORDS     HTG; CpG island; cysteine string; DnaJ; import receptor; KIAA1196;
             NG26; phosphoribulokinase; Prp1; Prp6; splicing factor; TOM;
             TPD52L2; tumor protein; Zer1.
SOURCE       Homo sapiens
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 139505)
AUTHORS      Almeida,J.
TITLE        Direct Submission
JOURNAL      Submitted (05-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Sep 19, 2000 this sequence version replaced gi:10129695.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases:
             Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
             was generated from part of bacterial clone contigs of human
             chromosome 20, constructed by the Sanger Centre Chromosome 20
             Mapping Group.  Further information can be found at
             http://www.sanger.ac.uk/HGP/Chr20
             IMPORTANT: This sequence is not the entire insert of clone
             RP4-591C20 It may be shorter because we sequence overlapping
             sections only once, except for a 100 base overlap.
             The true right end of clone RP4-591C20 is at 139505 in this
             sequence. The true left end of clone RP11-238J15 is at 42466 in
             this sequence. The true right end of clone RP5-824A14 is at 100 in
             this sequence. This sequence was finished as follows unless
             otherwise noted: all regions were either double-stranded or
             sequenced with an alternate chemistry or covered by high quality
             data (i.e., phred quality >= 30); an attempt was made to resolve
             all sequencing problems, such as compressions and repeats; all
             regions were covered by at least one plasmid subclone or more than
             one M13 subclone; and the assembly was confirmed by restriction
             digest. RP4-591C20 is from the library RPCI-4 constructed by the
             group of Pieter de Jong. For further details see
             http://www.chori.org/bacpac/home.htm
VECTOR:      pCYPAC2.
FEATURES             Location/Qualifiers
     source          1..139505
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP4-591C20"
                     /clone_lib="RPCI-4"
     repeat_region   696..747
                     /note="MER53 repeat: matches 1..55 of consensus"
     repeat_region   748..1063
                     /note="MER53 repeat: matches 1..313 of consensus"
     repeat_region   1064..1212
                     /note="AluSq repeat: matches 55..188 of consensus"
     repeat_region   2011..2060
                     /note="MER53 repeat: matches 1..53 of consensus"
     repeat_region   2061..2194
                     /note="AluJ/monomer repeat: matches 1..130 of consensus"
     misc_feature    2445..6791
                     /note="CpG island"
```

```
gene            /evidence=not_experimental
                2489..3898
                /gene="dJ591C20.1"
CDS             2489..3898
                /gene="dJ591C20.1"
                /note="match: cDNAs: Em:AK001207 Em:AL157426 Em:U95607
                Em:X57435
                match: ESTs: Em:AA664327 Em:AI207810 Em:AA779058
                match: proteins: Tr:O95870 Tr:Q92102 Tr:O69340 Tr:Q9VZF1
                Sw:P41879 Tr:Q69130 Tr:Q90783 Sw:P04258"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ591C20.1 (novel protein similar to mouse
                NG26)"
                /protein_id="CAC15491.1"
                /db_xref="GI:11125672"
                /translation="MCVICFVRALVRFVKIYLTASYTYPFRGWPVAFRWDDVRAVGRS
                SSHRALTCAAAAGVWLLRDETGLGGDALGRPPGARSOAQCLLQQLRELPGQLASYAL
                AHSLGRWLVYPGSVSLMTRALLPLLQQGQERLVERYHGRRAKLVACDGNEIDTMFMDR
                RQHPGSHVHGPRIVICCEGNAGFYEMGCLSAPLEAGYSVLGWNHPGFGSSTGVPFPQH
                DANAMDVVVEYALHRLHFPPAHLVVYGWSVGGFTATWATMTYPELGALVLDATFDDLV
                PLALKVMPHSWWKGLVVRTVREHFNLNVAEQLCCYPGPVLLRRTQDDVVSTSGRLRPL
                SPGDVEGNRGNELLLRLLEHRYPVVMAREGRAVVTRWLRAGSLAQEAAFYARYRVDED
                WCLALLRSYRARCEEELEGEEALGPHGPAFPWLVGQGLSSRRRRLALFLARKHLKNV
                EATHFSPLEPEEFQLPWRL"
repeat_region   5103..5232
                /note="MIR repeat: matches 9. .146 of consensus"
repeat_region   5430..5743
                /note="AluSq repeat: matches 6. .312 of consensus"
mRNA            join(6238..6332,10244..10389,14616..14764,16764..16823,
                23667..23768,30138..30186,30798..32493)
                /gene="TPD52L2"
                /note="match: cDNAs: Em:AF004430
                match: ESTs: Em:AW230830 Em:AA833100 Em:AA608309"
                /evidence=not_experimental
                /product="dJ591C20.2.1 (tumor protein D52-like 2)"
gene            6238..32493
                /gene="TPD52L2"
CDS             join(6314..6332,10244..10389,14616..14764,16764..16823,
                23667..23768,30138..30186,30798..30893)
                /gene="TPD52L2"
                /note="isoform 1
                match: proteins: Tr:O43399"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ591C20.2.1 (tumor protein D52-like 2)"
                /protein_id="CAC15492.1"
                /db_xref="GI:11125673"
                /translation="MDSAGQDINLNSPNKGLLSDSMTDVPVDTGVAARTPAVEGLTEA
                EEEELRAELTKVEEEIVTLRQVLAAKERHCGELKRRLGLSTLGELKQNLSRSWHDVQV
                SSAYVKTSEKLGEWNEKVTQSDLYKKTQETLSQAGQKTSAALSTVGSAISRKLGDMRN
                SATFKSFEDRVGTIKSKVVGDRENGSGNLPSSAGSGDKPLSDPAPF"
repeat_region   7042..7350
                /note="AluSc repeat: matches 1. .309 of consensus"
repeat_region   7399..7706
                /note="AluY repeat: matches 1. .309 of consensus"
repeat_region   7707..7893
                /note="AluJo repeat: matches 1. .311 of consensus"
repeat_region   7894..8201
                /note="AluSp repeat: matches 1. .308 of consensus"
repeat_region   8202..8216
                /note="AluJo repeat: matches 120. .134 of consensus"
repeat_region   8232..8530
                /note="AluSx repeat: matches 1. .300 of consensus"
repeat_region   8769..8885
                /note="AluSq/x repeat: matches 1. .117 of consensus"
repeat_region   8950..9159
                /note="AluSx repeat: matches 2. .212 of consensus"
repeat_region   9614..9896
                /note="AluSx repeat: matches 5. .292 of consensus"
misc_feature    complement(9791..10250)
                /note="match: GSS: Em:AQ888148"
misc_feature    complement(9810..10258)

repeat_region   10791..11090
                /note="match: GSS: Em:AQ583076"
repeat_region   11469..11666
                /note="AluJb repeat: matches 1. .298 of consensus"
repeat_region   11667..11963
                /note="AluY repeat: matches 106. .306 of consensus"
repeat_region   11989..12290
                /note="AluSg repeat: matches 1. .297 of consensus"
repeat_region   12302..12577
                /note="AluSx repeat: matches 1. .302 of consensus"
repeat_region   12587..12881
                /note="AluSx repeat: matches 21. .295 of consensus"
misc_feature    complement(12730..13211)
                /note="AluSq repeat: matches 1. .298 of consensus"
repeat_region   12986..13011
                /note="match: GSS: Em:AQ192102"
misc_feature    13231..13511
                /note="13 copies 2 mer tt 92% conserved"
                /gene="TPD52L2"
repeat_region   13322..13625
                /note="match: GSS: Em:AQ722319"
repeat_region   13639..13960
                /note="AluSx repeat: matches 1. .306 of consensus"
repeat_region   14156..14464
                /note="AluY repeat: matches 1. .311 of consensus"
repeat_region   14989..15297
                /note="AluSp repeat: matches 1. .309 of consensus"
repeat_region   15312..15622
                /note="AluSx repeat: matches 31. .311 of consensus"
repeat_region   15806..16101
                /note="AluJo repeat: matches 1. .312 of consensus"
repeat_region   16102..16127
                /note="AluSx repeat: matches 1. .294 of consensus"
repeat_region   17964..18272
                /note="13 copies 2 mer ta 96% conserved"
repeat_region   18273..18406
                /note="AluSp repeat: matches 1. .309 of consensus"
repeat_region   18776..18979
                /note="AluY repeat: matches 1. .137 of consensus"
repeat_region   19859..19957
                /note="12 copies 17 mer 61% conserved"
repeat_region   21557..21857
                /note="MIR repeat: matches 119. .225 of consensus"
misc_feature    21557..21769
                /note="AluY repeat: matches 1. .300 of consensus"
                /gene="TPD52L2"
misc_feature    2007..22329
                /gene="TPD52L2"


Query Match      36.0%;      Score 28.8;    DB 92;    Length 139505;
Best Local Similarity   60.0%;     Pred. No. 17;
Matches    48;    Conservative    0;    Mismatches    32;    Indels    0;    Gaps    0;

Qy      1 attcacatccaccagccatgagggtgcttgtactagctcttgctggctccgcagtgg 60
            | ||| || ||  |  |||||||| | |||||||| | || ||| | |||| ||
Db  76746 AATCAAATGCACACGCAGCGGAAGCTGCCTTGCTCCATGCACTGCCACTGGTTTCGTCCT 76805

Qy     61 gggaccagtccaacttgggg 80
            |||| |   ||| |||||||
Db  76806 GGAATCCGACCGTGTTCGGG 76825

RESULT 9
AC068717/c
LOCUS       AC068717    153713 bp    DNA    HTG    07-JUL-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-13K9, WORKING DRAFT SEQUENCE,
            13 unordered pieces.
ACCESSION   AC068717
VERSION     AC068717.2    GI:8468941
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
```

```
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
             Waterston,R.H.
TITLE        The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 153713)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (07-MAY-2000) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
COMMENT      On Jun 12, 2000 this sequence version replaced gi:7715643.
             --------------------- Genome Center -----------------------
             Center: Washington University.Genome Sequencing Center
             Center code: WUGSC
             Web site:http://genome.wustl.edu/gsc/index.shtml
             --------------- Project Information ----------------------
             Center project name: H_NH0013K09
             ----------------- Summary Statistics ----------------------
             Sequencing vector: M13; 100%
             Sequencing vector: plasmid; 0%
             Chemistry: Dye-primer ET; 100% of reads
             Chemistry: Dye-terminator Big Dye; 0% of reads
             Assembly program: Phrap; version 0.90319
             Consensus quality: 148042 bases at least Q40
             Consensus quality: 149465 bases at least Q30
             Consensus quality: 150268 bases at least Q20
             Insert size: 157000; agarose-fp
             Insert size: 152513; sum-of-contigs
             Quality coverage: 4.99 in Q20 bases; agarose-fp
             Quality coverage: 5.18 in Q20 bases; sum-of-contigs
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 13 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1    1595: contig of 1595 bp in length
        *     1596    1695: gap of unknown length
        *     1696    7734: contig of 6039 bp in length
        *     7735    7784: gap of unknown length
        *     7784   14154: contig of 6320 bp in length
        *    14155   14254: gap of unknown length
        *    14254   20602: contig of 6348 bp in length
        *    20602   20702: gap of unknown length
        *    20603   20702: contig of 9051 bp in length
        *    20703   29753: gap of unknown length
        *    29753   29853: contig of 7648 bp in length
        *    29854   37501: gap of unknown length
        *    37501   37601: contig of 6709 bp in length
        *    37602   44310: gap of unknown length
        *    44310   44410: contig of 8449 bp in length
        *    44311   44410: gap of unknown length
        *    44411   52859: contig of 7884 bp in length
        *    52860   52959: gap of unknown length
        *    52959   60843: contig of 14307 bp in length
        *    60843   60943: gap of unknown length
        *    60044   75250: contig of 14307 bp in length
        *    60944   75250: gap of unknown length
        *    75251   75350: contig of 18443 bp in length
        *    75350   93793: gap of unknown length
        *    93793  125972: contig of 32079 bp in length
        *    93894  125972: gap of unknown length
        *   125973  126072: contig of 27641 bp in length
        *   126073  153713: gap of unknown length.
FEATURES             Location/Qualifiers
     source          1..153713
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-13K9"
```

```
     misc_feature    1..1595
                     /note="assembly_name:Contig5
                     clone_end:T7
                     vector_side:left"
     misc_feature    1696..7734
                     /note="assembly_name:Contig6"
     misc_feature    7835..14154
                     /note="assembly_name:Contig7"
     misc_feature    14255..20602
                     /note="assembly_name:Contig8"
     misc_feature    20703..29753
                     /note="assembly_name:Contig9"
     misc_feature    29854..37501
                     /note="assembly_name:Contig10"
     misc_feature    37602..44310
                     /note="assembly_name:Contig11"
     misc_feature    44411..52859
                     /note="assembly_name:Contig12"
     misc_feature    52960..60843
                     /note="assembly_name:Contig13
                     clone_end:SP6
                     vector_side:left"
     misc_feature    60944..75250
                     /note="assembly_name:Contig14"
     misc_feature    75351..93793
                     /note="assembly_name:Contig15"
     misc_feature    93894..125972
                     /note="assembly_name:Contig16"
     misc_feature    126073..153713
                     /note="assembly_name:Contig17"
BASE COUNT    49043 a   26412 c   26181 g   50874 t   1203 others
ORIGIN
```

```
Query Match              35.8%;   Score 28.6;   DB 74;   Length 153713;
Best Local Similarity    67.8%;   Pred. No. 19;
Matches   40;  Conservative   0;  Mismatches   19;  Indels   0;  Gaps   0;

Qy      17 catgaggtgcttgtactagcttcttgctgtggtctccgcagtggggaccagtccaact 75
           | ||||||| |||||||| ||| ||  ||||| |||| || |||||||  ||||||||
Db   98480 CCTGAGGGAGGCATGCACCAGATCTTTCTCTAGTACTGCCAGATGAGGACCCAGCCAACT 98422
```

```
RESULT 10
AR070469
LOCUS        AR070469       10807 bp    DNA                        PAT  18-FEB-2000
DEFINITION   Sequence 5 from patent US 5905185.
ACCESSION    AR070469
VERSION      AR070469.1  GI:7221357
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1 (bases 1 to 10807)
  AUTHORS    Garner,I., Cottingham,I.R., Temperley,S.M., Foster,D.C.,
             Sprecher,C.A. and Prunkard,D.E.
TITLE        Protein C production in non-human transgenic mammals
JOURNAL      Patent: US 5905185-A 5 18-MAY-1999;
FEATURES             Location/Qualifiers
     source          1..10807
                     /organism="unknown"
BASE COUNT     2214 a    3066 c    2985 g    2542 t
ORIGIN
```

```
Query Match              35.5%;   Score 28.4;   DB 9;   Length 10807;
Best Local Similarity    62.9%;   Pred. No. 24;
Matches   44;  Conservative   0;  Mismatches   26;  Indels   0;  Gaps   0;

Qy       9 ccaccagccatgaggtgcttgtactagctctgtggctctcgcagtggggaccag 68
           ||  ||||||||||||||||| ||  | | ||||  | ||||  |||| ||||||
Db    4249 CCTGCAGCCATGAAGTGCCTCCTGCTGCCCTTGGCCCTGCTGCTGTGCGTGCCAG 4308
```

```
Qy        69 tccaacttgg 78
             ||| | |||
Db      4309 GCCATCATCG 4318


RESULT  11
147708
LOCUS          147708        10807 bp    DNA                  PAT       07-OCT-1997
DEFINITION     Sequence 7 from patent US 5639940.
ACCESSION      147708
VERSION        147708.1    GI:2471673
KEYWORDS       .
SOURCE         Unknown.
ORGANISM       Unknown.
               Unclassified.
REFERENCE      1  (bases 1 to 10807)
AUTHORS        Garner,I., Dalrymple,M.L., Prunkard,D.E. and Foster,D.C.
TITLE          Production of fibrinogen in transgenic animals
JOURNAL        Patent: US 5639940-A 7 17-JUN-1997;
FEATURES              Location/Qualifiers
     source        1..10807
                    /organism="unknown"
BASE COUNT     2214 a   3066 c   2985 g   2542 t
ORIGIN

Query Match          35.5%;   Score 28.4;  DB 10;  Length 10807;
Best Local Similarity 62.9%;  Pred. No. 24;
Matches   44;  Conservative   0;  Mismatches  26;  Indels   0;  Gaps   0;

Qy         9 ccaccagccataggggtgctgtactagctcttgctgtggctctcgcagtggggggaccag 68
             ||   |||||||||    ||  ||  || || ||| |  |||| ||||||    || ||||
Db      4249 CCTGCAGCCATGAGTGCCTCCTGCTTGCCCCTGGGCCTGGCCCTGCCTCGCCTGTGGGCGTCCAG 4308

Qy        69 tccaacttgg 78
             ||| | |||
Db      4309 GCCATCATCG 4318


RESULT  12
AC010920
LOCUS          AC010920     172815 bp    DNA                  HTG       15-FEB-2000
DEFINITION     Drosophila melanogaster chromosome X clone BACR47D09 (D895) RPCI-98
               47.D.9 map 14C-14D strain y; cn bw sp, *** SEQUENCING IN PROGRESS
               ***, 107 unordered pieces.
ACCESSION      AC010920
VERSION        AC010920.10  GI:6978364
KEYWORDS       HTG; HTGS_PHASE1.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
               Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
               Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1  (bases 1 to 172815)
AUTHORS        Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
               Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,N.L.,
               Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
               Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
               Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
               Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
               Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
               Wooley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
TITLE          Sequencing of Drosophila melanogaster
JOURNAL        Unpublished
REFERENCE      2  (bases 1 to 172815)
AUTHORS        Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
               Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
               Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
               Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
               Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
               Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
               Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
               Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
               Rubin,G.M.
TITLE
JOURNAL        Direct Submission
               Submitted (27-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
               Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT        On Feb 15, 2000 this sequence version replaced gi:6466918.
               For further information about this sequence, including its location
               and relationship to other sequences, please visit our sequence
               archive Web site (http://www.fruitfly.org/sequence/) or send email
               to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
               the following cutoffs: length >= 200 bases.
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 107 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as
               * runs of N, but the exact sizes of the gaps are unknown.
               * This record will be updated with the finished sequence
               * as soon as it is available and the accession number will
               * be preserved.
               *       582:  contig of 582 bp in length
               *       583:  gap of unknown length
               *       662:  contig of 1063 bp in length
               *       663:  gap of unknown length
               *      1725:  gap of unknown length
               *      1726:  gap of unknown length
               *      1805:  contig of 548 bp in length
               *      1806:  gap of unknown length
               *      2353:  gap of unknown length
               *      2354:  gap of unknown length
               *      2433:  gap of unknown length
               *      2434:  gap of unknown length
               *      3131:  contig of 698 bp in length
               *      3132:  gap of unknown length
               *      3211:  gap of unknown length
               *      3212:  gap of unknown length
               *      3838:  contig of 627 bp in length
               *      3839:  gap of unknown length
               *      3918:  gap of unknown length
               *      3919:  gap of unknown length
               *      4920:  contig of 1002 bp in length
               *      4921:  gap of unknown length
               *      5000:  gap of unknown length
               *      5001:  gap of unknown length
               *      5575:  contig of 575 bp in length
               *      5576:  gap of unknown length
               *      5655:  gap of unknown length
               *      5656:  gap of unknown length
               *      6243:  contig of 588 bp in length
               *      6244:  gap of unknown length
               *      6323:  gap of unknown length
               *      6324:  gap of unknown length
               *      6911:  contig of 588 bp in length
               *      6912:  gap of unknown length
               *      6991:  gap of unknown length
               *      6992:  gap of unknown length
               *      7531:  contig of 540 bp in length
               *      7532:  gap of unknown length
               *      7611:  gap of unknown length
               *      7612:  gap of unknown length
               *      8418:  contig of 807 bp in length
               *      8419:  gap of unknown length
               *      8498:  gap of unknown length
               *      8499:  gap of unknown length
               *      8826:  contig of 328 bp in length
               *      8827:  gap of unknown length
               *      8906:  gap of unknown length
               *      8907:  gap of unknown length
               *      9967:  contig of 1061 bp in length
               *      9968:  gap of unknown length
               *     10047:  gap of unknown length
               *     10048:  gap of unknown length
               *     11128:  contig of 1081 bp in length
               *     11129:  gap of unknown length
               *     11208:  gap of unknown length
               *     11209:  gap of unknown length
               *     12367:  contig of 1159 bp in length
               *     12368:  gap of unknown length
               *     12447:  gap of unknown length
               *     12448:  gap of unknown length
               *     13264:  contig of 817 bp in length
               *     13265:  gap of unknown length
               *     13344:  gap of unknown length
               *     13345:  gap of unknown length
               *     14345:  contig of 1001 bp in length
               *     14346:  gap of unknown length
               *     14425:  gap of unknown length
               *     14426:  gap of unknown length
               *     15339:  contig of 914 bp in length
               *     15340:  gap of unknown length
               *     15419:  gap of unknown length
               *     15420:  gap of unknown length
               *     16010:  contig of 591 bp in length
               *     16011:  gap of unknown length
               *     16090:  gap of unknown length
               *     16091:  gap of unknown length
               *     17621:  contig of 1531 bp in length
               *     17622:  gap of unknown length
               *     17701:  gap of unknown length
               *     17702:  gap of unknown length
               *     18720:  contig of 1019 bp in length
               *     18721:  gap of unknown length
               *     18800:  gap of unknown length
               *     18801:  gap of unknown length
               *     20314:  contig of 1514 bp in length
               *     20315:  gap of unknown length
               *     20394:  gap of unknown length
               *     20395:  gap of unknown length
               *     22174:  contig of 1780 bp in length
               *     22175:  gap of unknown length
               *     22254:  gap of unknown length
               *     22255:  gap of unknown length
               *     23537:  contig of 1283 bp in length
               *     23538:  gap of unknown length
               *     23617:  gap of unknown length
               *     23618:  gap of unknown length
               *     24301:  contig of 684 bp in length
               *     24302:  gap of unknown length
               *     24381:  gap of unknown length
               *     24382:  gap of unknown length
               *     25415:  contig of 1034 bp in length
               *     25416:  gap of unknown length
               *     25495:  gap of unknown length
               *     25496:  gap of unknown length
               *     26888:  contig of 1393 bp in length
               *     26889:  gap of unknown length
               *     26968:  gap of unknown length
               *     26969:  gap of unknown length
               *     27696:  contig of 728 bp in length
```

```
*  145946
   146199: contig of 254 bp in length
*  146200
   146279: gap of unknown length
*  146280
   146570: contig of 291 bp in length
*  146571
   146650: gap of unknown length
*  146651
   147165: contig of 515 bp in length
*  147166
   147245: gap of unknown length
*  147246
   148248: contig of 1003 bp in length
*  148249
   148328: gap of unknown length
*  148329
   149175: contig of 847 bp in length
*  149176
   149255: gap of unknown length
*  149256
   149664: contig of 409 bp in length
*  149665
   149744: gap of unknown length
*  149745
   150270: contig of 526 bp in length
*  150271
   150350: gap of unknown length
*  150351
   151005: contig of 655 bp in length
*  151006
   151085: gap of unknown length
*  151086
   151618: contig of 533 bp in length
*  151619
   151698: gap of unknown length
*  151699
   152268: contig of 570 bp in length
*  152269
   152348: gap of unknown length
*  152349
   153140: contig of 792 bp in length
*  153141
   153220: gap of unknown length


Query Match        35.5%;   Score 28.4;  DB 62;  Length 172815;
Best Local Similarity 66.1%;  Pred. No. 23;
Matches 41;  Conservative  0;  Mismatches 21;  Indels 0;  Gaps  0;

Qy        10 caccagccatgaggtgcttgctactagctcttgctgtcctccgcagtggggaccagt 69
             || |||||||||||| |  |  | || | | |  || |   |||  ||  |  |||||
Db    156970 CACGAGCCATGAGAGCTGTTCCTACTTACACTCTCCGTGGCCCTAGCCGTGGTCGCCGCCT 157029

Qy        70 cc 71
             ||
Db    157030 CC 157031


RESULT 13
S66686/c
LOCUS       S66686        840 bp    mRNA            INV       23-DEC-1993
DEFINITION  n-syb=vamp homolog [Drosophila, head, mRNA, 840 nt].
ACCESSION   S66686
VERSION     S66686.1  GI:436306
KEYWORDS    .
SOURCE      Drosophila sp. head.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 840)
  AUTHORS   DiAntonio,A., Burgess,R.W., Chin,A.C., Deitcher,D.L., Scheller,R.H.
            and Schwarz,T.L.
  TITLE     Identification and characterization of Drosophila genes for
            synaptic vesicle proteins
  JOURNAL   J. Neurosci. 13 (11), 4924-4935 (1993)
  MEDLINE   94045921
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 139319] from the original journal article.
            This sequence comes from Fig. 1.
            Map location: band 62A.
FEATURES             Location/Qualifiers
     source          1..840
                     /organism="Drosophila sp."
                     /db_xref="taxon:7242"
     gene            154..699
                     /gene="n-syb"
                     /note="vamp homolog"
     CDS             154..699
                     /gene="n-syb"
                     /note="synaptic vesicle protein; This sequence comes from
                     Fig. 1"
                     /codon_start=1
                     /protein_id="AAB28707.1"
                     /db_xref="GI:436307"
```

```
*  27697
   27776: gap of unknown length
*  27777
   29060: contig of 1284 bp in length
*  29061
   29140: gap of unknown length
*  29141
   30838: contig of 1698 bp in length
*  30839
   30918: gap of unknown length
*  30919
   32033: contig of 1115 bp in length
*  32034
   32113: gap of unknown length
*  32114
   33769: contig of 1656 bp in length
*  33770
   33849: gap of unknown length
*  33850
   36145: contig of 2296 bp in length
*  36146
   36225: gap of unknown length
*  36226
   37811: contig of 1586 bp in length
*  37812
   37891: gap of unknown length
*  37892
   40322: contig of 2431 bp in length
*  40323
   40402: gap of unknown length
*  40403
   42203: contig of 1801 bp in length
*  42204
   42283: gap of unknown length
*  42284
   44168: contig of 1885 bp in length
*  44169
   44248: gap of unknown length
*  44249
   46623: contig of 2375 bp in length
*  46624
   46703: gap of unknown length
*  46704
   50156: contig of 3453 bp in length
*  50157
   50236: gap of unknown length
*  50237
   53444: contig of 3208 bp in length
*  53445
   53524: gap of unknown length
*  53525
   56137: contig of 2613 bp in length
*  56138
   56217: gap of unknown length
*  56218
   60505: contig of 4288 bp in length
*  60506
   60585: gap of unknown length
*  60586
   64288: contig of 3703 bp in length
*  64289
   64368: gap of unknown length
*  64369
   68544: contig of 4176 bp in length
*  68545
   68624: gap of unknown length
*  68625
   73051: contig of 4427 bp in length
*  73052
   73131: gap of unknown length
*  73132
   78613: contig of 5482 bp in length
*  78614
   78693: gap of unknown length
*  78694
   84731: contig of 6038 bp in length
*  84732
   84811: gap of unknown length
*  84812
   90313: contig of 5502 bp in length
*  90314
   90393: gap of unknown length
*  90394
   99366: contig of 8973 bp in length
*  99367
   99446: gap of unknown length
*  99447
   108532: contig of 9086 bp in length
*  108533
   108612: gap of unknown length
*  108613
   120831: contig of 12219 bp in length
*  120832
   120911: gap of unknown length
*  120912
   138139: contig of 17228 bp in length
*  138140
   138219: gap of unknown length
*  138220
   138731: contig of 512 bp in length
*  138732
   138811: gap of unknown length
*  138812
   139262: contig of 451 bp in length
*  139263
   139342: gap of unknown length
*  139343
   139741: contig of 399 bp in length
*  139742
   139821: gap of unknown length
*  139822
   140293: contig of 472 bp in length
*  140294
   140373: gap of unknown length
*  140374
   141021: contig of 648 bp in length
*  141022
   141101: gap of unknown length
*  141102
   141858: contig of 757 bp in length
*  141859
   141938: gap of unknown length
*  141939
   142495: contig of 557 bp in length
*  142496
   142575: gap of unknown length
*  142576
   143136: contig of 561 bp in length
*  143137
   143216: gap of unknown length
*  143217
   143690: contig of 474 bp in length
*  143691
   143770: gap of unknown length
*  143771
   144232: contig of 462 bp in length
*  144233
   144312: gap of unknown length
*  144313
   144924: contig of 612 bp in length
*  144925
   145004: gap of unknown length
*  145005
   145865: contig of 861 bp in length
*  145866
   145945: gap of unknown length
```

```
                          /translation="MADAAPAGDAPPNAGAPAGEGDGEIVGGPHNPQQIAAQKRLQQ
                          TQAQVDEVVDIMRTNVEKVLERDTKLSELDDRADALQQGASQFEQOAGKLKRKFWLQN
                          LKMMIIMGVIGLVVVGIIANKLGLIGGEQPPYOQYPPQYMOPPPPPPPQQPAGGQSSLV
                          DGAGAGDGAGGSAGAGDHGGV"
BASE COUNT      234 a      230 c      257 g      119 t
ORIGIN

Query Match
Best Local Similarity    35.2%;    Score 28.2;    DB 6;    Length 840;
Matches     39;   Conservative   0;   Mismatches   18;   Indels   0;   Gaps   0;

Qy     4  cacatccaccagccatgagggtgcttgtactagctcttgctgtggctctcgcagtgg  60
          | ||||||||| |||| || ||| || ||| ||| ||||| |||||  || ||||
Db   639  CCCATCCACCAGCCGATGACTGTCCTCCGCTGGCTGCTGCGGTGGCGGGCGGCGGTGG  583

RESULT 14
AC01823i/c
LOCUS       AC018231     50701 bp    DNA                 HTG       09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC018231
VERSION     AC018231.1   GI:6552960
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 50701)
AUTHORS     Adams,M. and Venter,J.C.
TITLE       Direct Submission
JOURNAL     Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville. MD, USA
COMMENT     This sequence was identified as CDM:10213948 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES         Location/Qualifiers
      source     1..50701
                 /organism="Drosophila melanogaster"
                 /db_xref="taxon:7227"
BASE COUNT    13744 a    12081 c    12135 g    12741 t
ORIGIN

Query Match
Best Local Similarity    35.2%;    Score 28.2;    DB 65;    Length 50701;
Matches     39;   Conservative   0;   Mismatches   18;   Indels   0;   Gaps   0;

Qy     4  cacatccaccagccatgagggtgcttgtactagctcttgctgtggctctcgcagtgg  60
          | ||||||||| |||| || ||| || ||| ||| ||||| |||||  || ||||
Db   795  CGCATCCACCAGCGATGACTGTCCTCCGCTGGCTGGCTGCTGCGGTGTGGCGGGCGGCGGTGG  739

RESULT 15
SYCSLRB/c
LOCUS       SYCSLRB     146271 bp    DNA                 BCT       13-FEB-1999
DEFINITION  Synechocystis sp. PCC6803 complete genome, 19/27, 2392729-2538999.
ACCESSION   D64000 AB001339
VERSION     D64000.1   GI:1001484
KEYWORDS    RadC; 16S rRNA; 23S rRNA; 3-hydroxyisobutyrate dehydrogenase; 30S
            ribosomal protein S21; 50S ribosomal protein L20; 50S ribosomal
            protein L35; 5S rRNA; ABC transporter; ABC1-like; ATP-dependent Clp
            protease regulatory subunit; DNA recombinase; GMP synthetase; Mg2+
            transport ATPase; RadC protein; UDP-3-o-[3-hydroxymyristoyl]
            glucosamine n-acyltransferase; UDP-N-acetylglucosamine
            l-carboxyvinyltransferase; amidophoribosyltransferase;
            aspartate aminotransferase; carboxyl-terminal processing protease;
            cell division protein FtsH; cobinamide kinase; cytochrome b6;
```

```
            cytochrome b6-f complex subunit 4; cytosine-specific
            methyltransferase; ferredoxin; fructose-1,6-bisphosphate aldolase;
            fumarase; glucose inhibited division protein A; glucose transport
            protein; glucose-1-phosphate thymidylyltransferase;
            glutamate-1-semialdehyde 2,1- aminomutase; glycyl-tRNA synthetase;
            beta chain; high-affinity branched-chain amino acid transport
            ATP-binding protein; histidyl-tRNA synthetase; lipid A disaccharide
            synthase; mannosyl transferase; membrane-bound lytic
            transglycosylase A; methylenetetrahydrofolate dehydrogenase;
            oligopeptide transport system permease protein; penicillin-binding
            protein 1A; photosystem II PsbT protein; plastocyanin; protease IV;
            protein-export membrane protein SecD; protein-export membrane
            protein SecF; protochlorophyllide reductase ChlB subunit;
            ribonuclease D; ribonuclease III; ribulose bisphosphate carboxylase
            large subunit; ribulose bisphosphate carboxylase small subunit;
            sensory transduction histidine kinase; tRNA-Ala; tRNA-Gly;
            tRNA-Ile; tRNA-Ser; thiol-specific antioxidant protein; thioredoxin
            M; transcriptional regulatory protein HypF; transposase;
            triosephosphate isomerase; unknown function; Ycf33; Ycf38 gene.
SOURCE      Synechocystis sp. (strain:PCC6803) DNA.
  ORGANISM  Synechocystis sp.
            Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
REFERENCE   1 (bases 1 to 146271)
AUTHORS     Kaneko,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N.,
            Sugiura,M. and Tabata,S.
TITLE       Sequence analysis of the genome of the unicellular cyanobacterium
            Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
            region from map positions 64% to 92% of the genome
JOURNAL     DNA Res. 2 (4), 153-166 (1995)
MEDLINE     96127529
REFERENCE   2 (bases 1 to 146271)
AUTHORS     Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y.,
            Miyajima,N., Hirosawa,M., Sugiura,M., Sasamoto,S., Kimura,T.,
            Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K.,
            Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A.,
            Yamada,M., Yasuda,M. and Tabata,S.
TITLE       Sequence analysis of the genome of the unicellular cyanobacterium
            Synechocystis sp. strain PCC6803. II. Sequence determination of the
            entire genome and assignment of potential protein-coding regions
JOURNAL     DNA Res. 3 (3), 109-136 (1996)
MEDLINE     97061201
REFERENCE   3 (bases 1 to 146271)
AUTHORS     Tabata,S.
TITLE       Direct Submission
JOURNAL     Submitted (30-AUG-1995) to the DDBJ/EMBL/GenBank databases. Satoshi
            Tabata, Kazusa DNA Research Institute, Gene Structure 2; 1532-3
            Yanauchino, Kisarazu, Chiba 292, Japan (E-mail:tabata@kazusa.or.jp,
            Tel:0438-52-3933, Fax:0438-52-3934)
COMMENT     Potential protein coding regions were assigned on the basis of
            similarity search of the ORFs and GeneMark analysis.
FEATURES         Location/Qualifiers
    source    1..146271
                 /organism="Synechocystis sp."
                 /strain="PCC6803"
                 /db_xref="taxon:1143"
    CDS       1..675
                 /note="ORF_ID:slr0388"
                 /codon_start=1
                 /transl_table=11
                 /product="hypothetical protein"
                 /protein_id="BAA1110.1"
                 /db_xref=GI:1001485"
                 /translation="MWQSIRSLIFQTPCPLCQKPSQDLCRDCTGQILSHRFPRKFQQ
            WQGNFPRFVWGQYEGQLRRALTVMKFEQQPDIGITWLGEQLAEQWLQQPQAKTKLRPQV
            VPIPLNARKQAQRGFNQAERIAAGFCRLTGYVLHPQALQRVKDTQALFGLSPGDRRRE
            LQSALATGPGFNTRRPWLILDDIITTGTTALEARRIMEEKGAKVLGIVAIAAPSFSQP
            MELRQQ"
    CDS       765..1379
                 /note="ORF_ID:slr0769"
                 /codon_start=1
                 /transl_table=11
                 /product="hypothetical protein"
                 /protein_id="BAA10111.1"
```

CDS
```
/db_xref="GI:1001486"
/translation="MQKQVSPLARVALMLGGVSLTATSLVIPTAAQAQSSLNGGSAIP
PLKPIPESSTYTPPAAPSTPSNPTSQALATHIVLVLGERKVTAYQNDKVLASYPVAVG
KKGWETPQGNFQVIQMVENPWENPWNGKKVAASLDGFPIGIRWIGFWSDGKNTIGFHG
TPKKHEHFLGTAASHGCVRMRNQDVVALFNMVQNGTPVRVQKK"
1504..1824
/note="ORF_ID:slr0770"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA10112.1"
/db_xref="GI:1001487"
/translation="MATTLINHIEITPGTCGGKPRIAGHRITVQDIVIWHERMGLSPD
EIVHHYPSITLADVYAALSYYHDHLEEIRQGIIDDEIFAEENRSSTPSLMEAKLRKVH
GKQD"
1808..2143
/note="ORF_ID:slr0771"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA10113.1"
/db_xref="GI:1001488"
/translation="MANRIKFHLDENISYSIALGLRHRSLDVTTTPETGLLGVSDQEQ
LQFAGSEGRVIFTQDTDFLKLAQSSPDHTGIAYCSQRNKSIDQVIQGLVLMWEILEPE
EMCNHIEFL"
```

gene
```
2375..3901
/gene="chlB"
```

CDS
```
2375..3901
/gene="chlB"
/note="ORF_ID:slr0772"
/codon_start=1
/transl_table=11
/product="protochlorophyllide reductase ChlB subunit"
/protein_id="BAA10114.1"
/db_xref="GI:1001489"
/translation="MKLAYWMYAGPAHIGTLRIASSFKNVHAIMHAPLGDDYFNVMRS
MLEREQNFTPVTTSVVDRQVLSRGSGEKVGTVIVRKDGEETPDLIVLPTCTSSILQE
DLANFVDRAQMDAHCDVLLADVNHYRYNELQAGDRTLKQIVEFYIKKARKQGNLATEK
TAKPSVNLIGFTTLGFHNQHDCTELKRLMADLGIEVNLLPEKATVDQLAKIPQAWFN
LCPYREIGLMTAEYLQEEFAQPYVDITPMGVVETARCIRKMQQVLNQQGFDVNYEDFI
QQQTRHVSQAAWFSRSIDCQNLTGKKAVVFGDNTHAAAMTKILAREMGIHVVLAGTYC
KYDADWFRAEVSEYCDEVLISEDNGAIADAIARIEPAAIFGTQMERHVGKRLDIPCGV
IAAPIHIQNFPLGYKPFLGYEGTNQIADLVYNSPTLGMEDHLLEIFGGHDTKEVITKG
ISSDSLGWHSTAQAELNKVPGFVRGKVKRNTEKFARERGLSEITLEWMYAAKEAVGA"
4033..4713
/note="ORF_ID:slr0773"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA10115.1"
/db_xref="GI:1001490"
/translation="MYILIGGMGSMGSNLAKNLLKMGHTVAAVDTDPLACQYAREKIG
VMAFEGSAVNTSVLLEAGVRKAGAVIAALQEDALNLALVTLSKHYGVPQIVVRMSDGD
FAEPYRLAGATHIISTTEALNRVINAIEYPQVDAMMHFEQGQIEVLKLSIPPNCTIV
GRSVAAIAQDPRFPSGTLIIGYQANAHEDLKIPNGNTLLEDGSTILAVTKPELVRQLI
DFMGLCQS"
```

CDS
```
complement(4750..6390)
/note="ORF_ID:sll0772"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA10116.1"
/db_xref="GI:1001491"
/translation="MGLICGLAIASGLTVTIPSEAQTISVNEFFPPVNRYQLGASNLG
NGVNQVTSVAELRDVQPSDWAFAALQSLVERYGCLVGYPDRTYRGDAEGTLRARPLSR
YEFAAGLNACLNTIEQLLQENVSVAQGDLDLLKKLLAQDFQAELKQLAVRVDNLETRTA
FLEDHQFSVTVKLYGQTITSLDKVFGDRVGGGSNEFQTQFAYRVRFNLETSFTGKDLF
RTRLQFSNFLNGVEQTGTNWTRFNYDDNSNNNVEVSHLWYRTPLTDNLLRLGTVGVG
YTDLVDTLTPPTIADDALGIPSRFGEYDPVYRRGGGAGFNWQITPTVQLSVGYLATN
PNPDPDPGNGLFNGGYHTLAQLAYQTADGGIGFTYSRSYFPAGNTDLMAGTGSLLAIQP
FGEVIATAGNFYTLQGYYRITPHVOIHGWGGFVEAQAQGGLSNLSNGVCGTVLRDVS.
WNSSAIWYGLVGISFPDVGGEGNLPGIALGIPPTVTASNLPGAVGQTTPYHLEAFYRI
```

```
QLNDNISITPGFWVLNPEANSNNATQYVGHIRTSFLF"
complement(6482..7888)
/gene="glcP"
```

gene
```
complement(6482..7888)
/gene="glcP"
```

CDS
```
/note="ORF_ID:sll0771"
/codon_start=1
/transl_table=11
/product="glucose transport protein"
/protein_id="BAA10117.1"
/db_xref="GI:1001492"
/translation="MNPSSSPSQSTANVKFVLLISGVAALGFLFGPTAVINGAVAA
LQKHFQTDSLLTGLSVSLALLGSALGAFGAGPIADRHGRIKTMILAAVLFTLSSIGSG
LPFTIWDFIEWRVLGGIGVGAASVIAPYIAEVSPAHLRGRLGSLQQLAIVSGIFIAL
LSNWFIALMAGGSAQNPWLFGAAAWRKMFWTELIPALLYGVCAFLIPESPRYLVAQGQ
GEKAAAILWKVEGGDVPSRIEEIQATVSLDHKPRFSDLLSRGGLLPIVWIGMGLSAL
QQFVGINVIFYYSSVLWRSVGFTEEKSLLITVTGFINLTTLVAIAFVDKFWFGRKPLL
LMGSIGMTITLGILSVVFGGATVVNGOPTLTGAAGIIALVTANLYVFSFGFSWGPIVW
VLLGEMFNNKIRAAALSVAAGVQWIANFIISTFPPLLDTVGLGPAYGLYATSAAISI
FFIWFFVKETKGKTLEQM"
8186..9604
/gene="secD"
```

gene
```
8186..9604
/gene="secD"
```

CDS
```
/note="ORF_ID:slr0774"
/codon_start=1
/transl_table=11
/product="protein-export membrane protein SecD"
/protein_id="BAA10118.1"
/db_xref="GI:1001493"
/translation="MQRLRWLLLLIVVLVVIGASFVLVKLPLQLGLDLRGGAQLTIEVQ
PTKEIPQIDNDSLVAVKTVIENRVNALGVSEPLVQTAGEDKIVQLPGVTDPGQAERI
LGGTAQLEFQQQRPGTEGEFQAEYSIKRQLDAELENLRRSGASPENSDRLEELIKAKE
ESNKALALFEPMGLTGKNLTDARPSPNQSGTAWEVALRFDEEGGQKFAELTQAVAGT
GRSLGVFLDNDLISAPVVGVEFANTGITGGAAVITGNFTIDTANDLAVQLRGGSLPFP
VEVVENRTVGATLGQESIRRSLVAGFVGLVLVLVFMAVYYRLPGIVADISLMIYAVLT
LAAFALVGVTLTLPGIAGFILSIGMAVDANVLIFERTEELRAGNTLYRSVEAGFFRA
FSSILDSNVTTLIACAALFWFGSGLVKGFALTLAIGVMVSLFTALTCSRTLLLVIVLS
LPKVRQNPRLFCPNLSSVTAKS"
9665..10612
/gene="secF"
```

gene

```
Query Match            35.2%;   Score 28.2;   DB  3;   Length 146271;
Best Local Similarity  64.6%;   Pred. No. 27;
Matches    42;   Conservative    0;   Mismatches   23;   Indels    0;   Gaps    0;

Qy     8 tccaccagccatgaggtgcttgtactagctcttgctgtgcctccgcagtggggacca 67
          ||| | |||| |  ||||||| | ||||||| |  ||||| |  ||||| ||
Db 75902 TCCCCGGGCCAACGCGGGTGCTGGAACTTTCCCCTTGAGGAGGCTCCGCAACTGGGTCACAA 75843

Qy    68 gtcca 72
          |  ||
Db 75842 TTACA 75838

Search completed: May  5, 2001, 22:04:19
Job time: 2423 sec
```

```
              GenCore version 4.5
        Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May  3, 2001, 15:51:41 ; Search time 12.84 Seconds
                                    (without alignments)
                                    112.397 Million cell updates/sec

Title:         US-09-426-776-10
Perfect score: 96
Sequence:      1 MRVLVLALAVALAVGDGSNLG 21

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters:    198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     PIR_67:*
               1: pir1:*
               2: pir2:*
               3: pir3:*
               4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	82	85.4	1788	2	T31095	vitellogenin precu
2	53	55.2	678	2	S46963	exeD protein - Aer
3	53	55.2	1687	2	T43144	vitellogenin II pr
4	45	46.9	330	2	T25169	hypothetical prote
5	45	46.9	691	2	T41724	probable adenosine
6	44	45.8	328	2	H75251	glycerol-3-phospha
7	44	45.8	585	2	S06958	sphingomyelin phos
8	44	45.8	629	1	A39825	sphingomyelin phos
9	44	45.8	992	2	S54396	protein L precurso
10	43	44.8	127	2	B75353	hypothetical prote
11	43	44.8	222	2	JC1384	beta-casein precur
12	43	44.8	273	2	C70551	hypothetical precur
13	43	44.8	346	2	H82637	glycerol-3-phospha
14	43	44.8	359	2	T44816	brp protein [impor
15	43	44.8	583	2	H75561	ABC transporter, A
16	43	44.8	598	2	T42070	protein serine/thr
17	43	44.8	683	2	I52996	transforming growt
18	43	44.8	729	2	T46328	probable adenosine
19	42.5	44.3	80	2	A65210	hypothetical 7.4 k
20	42.5	44.3	310	2	C83125	hypothetical prote
21	42.5	44.3	893	2	T03864	hypothetical prote
22	42	43.8	168	2	A31770	pheromone-binding
23	42	43.8	174	2	A64442	hypothetical prote
24	42	43.8	258	2	JH0472	apolipoprotein A-I
25	42	43.8	297	2	D69404	hypothetical prote
26	42	43.8	378	2	D64181	probable cytochrom
27	42	43.8	460	2	C70576	probable mgtE prot
28	42	43.8	550	2	C64867	probable sulfate t
29	42	43.8	654	2	H65125	probable general s
30	41.5	43.2	351	2	B81721	cytochrome D ubiqu
31	41	42.7	109	2	D75506	hypothetical prote
32	41	42.7	180	1	LGGT	beta-lactoglobulin
33	41	42.7	180	1	LGSH	beta-lactoglobulin
34	41	42.7	222	2	A32979	beta-casein precur
35	41	42.7	224	1	KBBOA2	beta-casein precur
36	41	42.7	226	1	KBHU	beta-casein precur
37	41	42.7	228	2	JT0564	beta-casein precur
38	41	42.7	231	2	F69147	UDP-N-acetylglucos
39	41	42.7	247	2	G82956	hypothetical prote
40	41	42.7	312	2	S18387	heme oxygenase - r
41	41	42.7	345	2	S55377	urPAB protein prec
42	41	42.7	387	2	A53586	albumin-binding pr
43	41	42.7	397	2	B70763	probable membrane
44	41	42.7	429	2	C83025	probable phosphose
45	41	42.7	597	2	C69283	hypothetical prote

ALIGNMENTS

```
RESULT    1
T31095
vitellogenin precursor - Oreochromis aureus
C;Species: Oreochromis aureus
C;Date:.19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C;Accession: T31095
R;Lim, E.H.; Lam, T.J.; Ding, J.L.
submitted to the EMBL Data Library, August 1997
A;Description: Cloning of full-length Oreochromis aureus vitellogenin cDNA and its
A;Reference number: 220978
A;Accession: T31095
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1788 <LIM>
A;Cross-references: EMBL:AF017250; NID:g4102880; PIDN:g4102881; PIDN:AAD01615.1
C;Superfamily: vitellogenin

Query Match             85.4%;   Score 82;  DB 2;  Length 1788;
Best Local Similarity   95.0%;          Pred. No. 0.00062;
Matches   19;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 MRVLVLALAVALAVGDGSNL 20
          |||||||||||||||| |||
Db      1 MRVLVLALAVALAVGDQSNL 20


RESULT    2
S46963
exeD protein - Aeromonas salmonicida
C;Species: Aeromonas salmonicida
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
R;Karlyshev, A.V.; MacIntyre, S.
Gene 158, 77-82, 1995
A;Title: Cloning and study of the genetic organization of the exe gene cluster of
A;Reference number: I39675; MUID:95309729
A;Accession: I39678
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-678 <RES>
A;Cross-references: EMBL:X80505; NID:g516349; PIDN:CAA56668.1; PID:g516351
C;Genetics:
A;Gene: exeD

Query Match             55.2%;   Score 53;  DB 2;  Length 678;
Best Local Similarity   60.0%;          Pred. No. 4.1;
Matches   12;  Conservative   3;  Mismatches   5;  Indels   0;  Gaps   0;

Qy      2 RVLVLALAVALAVGDGSNLG 21
          :||| |: : :|.||| |||
```

Db   352 QVLVEAIIVEIADGDGLNLG 371

RESULT 3
T43144
vitellogenin.II precursor - mummichog
C;Species: Fundulus heteroclitus (mummichog)
C;Date: 11-Jan-2000 #sequence_revision 18-Feb-2000
C;Accession: T43144
R;LaFleur, G.J.; Hoch, K.L.; Denslow, N.; Byrne, B.M.; Wallace, R.A.
submitted to the EMBL Data Library, September 1996
A;Description: Derivation of oocyte and egg proteins from parental vitellogenins in Fund
n.
A;Reference number: Z22316
A;Accession: T43144
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1687 <LAF>
A;Cross-references: EMBL:U70826; NID:g1621358; PID:g1621359; PIDN:AAB17152.1
C;Experimental source: estrogen-induced liver
C;Superfamily: vitellogenin
C;Keywords: egg yolk; phosphoprotein
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-1687/Product: vitellogenin II #status predicted <MAT>

Query Match   55.2%; Score 53; DB 2; Length 1687;
Best Local Similarity   75.0%; Pred. No. 8.7;
Matches   12; Conservative   1; Mismatches   3; Indels   0; Gaps   0;

Qy   1 MRVLVLALAVALAVGD 16
          || ||||||| ||| |:
Db   1 MRVLVLALTVALVAGN 16

RESULT 4
T25169
hypothetical protein T23F1.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T25169
R;Wilkinson, J.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19990
A;Accession: T25169
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-330 <WIL>
A;Cross-references: EMBL:Z81129; PIDN:CAB03405.1; GSPDB:GN00023; CESP:T23F1.6
C;Genetics:
A;Gene: CESP:T23F1.6
A;Map position: 5
A;Introns: 16/3
C;Superfamily: gliadin

Query Match   46.9%; Score 45; DB 2; Length 330;
Best Local Similarity   45.0%; Pred. No. 31;
Matches   9; Conservative   5; Mismatches   6; Indels   0; Gaps   0;

Qy   1 MRVLVLALALAVGDGSNL 20
          || ::: ||||||| |::
Db   1 MRCVILFAVALAVAQASSI 20

RESULT 5
T41724
probable adenosinetriphosphatase (EC 3.6.1.3) SPAC821.13c [similarity] - fission yeast (
C;Species: Schizosaccharomyces pombe
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C;Accession: T41724
R;Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, October 1999
A;Reference number: Z22012
A;Accession: T41724
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-691 <RIE>
A;Cross-references: EMBL:AL121770; PIDN:CAB57447.1; GSPDB:GN00066; SPDB:SPAC821.13c
A;Experimental source: strain 972h-; cosmid c821
C;Genetics:
A;Gene: SPDB:SPAC821.13c
A;Map position: 1
C;Keywords: hydrolase

Query Match   46.9%; Score 45; DB 2; Length 691;
Best Local Similarity   64.3%; Pred. No. 58;
Matches   9; Conservative   2; Mismatches   3; Indels   0; Gaps   0;

Qy   6 LALAVALAVGDGSN 19
          | | ||:|||:||:|
Db   362 LEKAVTLAIGDGAN 375

RESULT 6
H75251
glycerol-3-phosphate dehydrogenase, NAD(+)-dependent - Deinococcus radiodurans (stra:
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: H75251
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.:
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: H75251
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-328 <WHI>
A;Cross-references: GB:AE002091; GB:AE000513; NID:g6460446; PIDN:AAF12158.1; PID:g64(
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2621
A;Map position: 1
C;Superfamily: glycerol-3-phosphate dehydrogenase

Query Match   45.8%; Score 44; DB 2; Length 328;
Best Local Similarity   52.6%; Pred. No. 44;
Matches   10; Conservative   3; Mismatches   6; Indels   0; Gaps   0;

Qy   3 VLVLALAVALAVGDGSNLG 21
          || :||| :|||:||| ||
Db   184 VLKNVIAVAAGMGDGLHLG 202

RESULT 7
S06958
sphingomyelin phosphodiesterase, acidic, inactive splice form 2 - human
N;Alternate names: acid sphingomyelinase, splice form 2
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1993 #sequence_revision 24-Nov-1999 #text_change 21-Jan-2000
C;Accession: S06958; A39825
R;Quintern, L.E.; Schuchman, E.H.; Levran, O.; Suchi, M.; Ferlinz, K.; Reinke, H.; S:
EMBO J. 8, 2469-2473, 1989
A;Title: Isolation of cDNA clones encoding human acid sphingomyelinase: occurrence o:
A;Reference number: S06957; MUID:90060003
A;Accession: S06958
A;Molecule type: mRNA
A;Residues: 270-585 <QUI>
A;Cross-references: EMBL:X52679; NID:g28881; PIDN:CAA36902.1; PID:g825629
R;Schuchman, E.H.; Suchi, M.; Takahashi, T.; Sandhoff, K.; Desnick, R.J.
J. Biol. Chem. 266, 8531-8539, 1991

```
A;Title: Human acid sphingomyelinase. Isolation, nucleotide sequence, and expression of
A;Reference number: A39825
A;Accession: A39825
A;Molecule type: mRNA
A;Residues: 1-362,'TGGFYALSPYPCGLRLISLNMNFCSRENFWLL','INSTDPAGQLQWLVGELQAAEDRGD',375-585
A;Cross-references: GB:M59916; NID:g179094; PIDN:AAA58377.1; PID:g179095
C;Comment: This form lacked catalytic activity when expressed in COS-1 cells. Its functi
C;Comment: For the active splice form 1, see PIR:A39825.
C;Genetics:
A;Gene: GDB:SMPD1
A;Cross-references: GDB:128144; OMIM:257200
A;Map position: 11p15.4-11p15.4
A;Introns: 104/3; 362/2; 375/3; 401/2; 450/1
C;Superfamily: acid sphingomyelinase; phosphoesterase core homology; saposin repeat homo
C;Keywords: alternative splicing
F;81-175/Domain: saposin repeat homology <SAP>
F;200-320/Domain: phosphoesterase core homology <PEC>


Query Match                45.8%;    Score 44;  DB 2;  Length 585;
Best Local Similarity      71.4%;    Pred. No. 71;
Matches      10;  Conservative      2;  Mismatches      2;  Indels      0;  Gaps      0;

Qy      3  VLVLALAVALAVGD 16
           || |||:|||:| |
Db     36  VLALALALALSD 49


RESULT    8
A39825
sphingomyelin phosphodiesterase (EC 3.1.4.12), acidic, splice form 1 precursor - human
N;Alternate names: acid sphingomyelinase
C;Species: Homo sapiens (man)
C;Date: 20-Mar-1997 #sequence_revision 07-Jul-1995 #text_change 24-Nov-1999
C;Accession: S27009; S36357; A39825; JX0276; S06957; I55567; A42801; B42801; C42801; D42
R;Newrzella, D.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 373, 1233-1238, 1992
A;Title: Molecular cloning of the acid sphingomyelinase of the mouse and the organizatio
A;Reference number: S27009; MUID:93183402
A;Accession: S27009
A;Molecule type: DNA
A;Residues: 1-629 <NEW>
A;Cross-references: EMBL:X63600; NID:g556808
R;Hofmann, K.
submitted to the EMBL Data Library, December 1991
A;Reference number: S36357
A;Accession: S36357
A;Molecule type: DNA
A;Residues: 1-321,'T',323-629 <HOF>
A;Cross-references: EMBL:X63600; NID:g556808; PIDN:CAA45145.1; PID:g556809
R;Schuchman, E.H.; Suchi, M.; Takahashi, T.; Sandhoff, K.; Desnick, R.J.
J. Biol. Chem. 266, 8531-8539, 1991
A;Title: Human acid sphingomyelinase. Isolation, nucleotide sequence, and expression of
A;Reference number: A39825; MUID:91217097
A;Accession: JX0276
A;Molecule type: mRNA
A;Residues: 1-629 <SCH>
A;Cross-references: GB:M59916; NID:g179094; PIDN:AAA58377.1; PID:g179095
A;Note: polymorphisms were demonstrated at positions 322 and 506
R;Ida, H.; Rennert, O.M.; Eto, Y.; Chan, W.Y.
J. Biochem. 114, 15-20, 1993
A;Title: Cloning of a human acid sphingomyelinase cDNA with a new mutation that renders
A;Reference number: JX0276; MUID:94012573
A;Accession: JX0276
A;Molecule type: mRNA
A;Residues: 1-35,38-156,'R',158-321,'T',323-505,'G',507-629 <IDA>
A;Cross-references: GB:M59916; NID:g179094; PIDN:AAA58377.1; PID:g402621
A;Note: polymorphisms were demonstrated at positions 322 and 506; PIDN:CAA42584.1; PID:g402621
A;Note: the authors translated the codon CTC for residue 85 as Thr and ACC for residue 8
A;Note: this form has no catalytic activity; correction of position 157 to Cys restores
R;Quintern, L.E.; Schuchman, E.H.; Levran, O.; Suchi, M.; Ferlinz, K.; Reinke, H.; Sandh
EMBO J. 8, 2469-2473, 1989
A;Title: Isolation of cDNA clones encoding human acid sphingomyelinase: occurrence of al
A;Reference number: S06957; MUID:90060003
```

```
A;Accession: S06957
A;Molecule type: mRNA
A;Residues: 128-629 <QUI>
A;Cross-references: EMBL:X52678; NID:g28879; PIDN:CAA36901.1; PID:g28880
A;Note: parts of this sequence were confirmed by peptide sequencing
R;Levran, O.; Desnick, R.J.; Schuchman, E.H.
J. Clin. Invest. 88, 806-810, 1991
A;Title: Niemann-Pick type B disease. Identification of a single codon deletion in
A;Reference number: I55567; MUID:91358737
A;Accession: I55567
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 119-120,'H',122 <LEV>
A;Cross-references: GB:S55766; NID:g234719; PIDN:AAB19680.1; PID:g234720
R;Takahashi, T.; Suchi, M.; Desnick, R.J.; Takada, G.; Schuchman, E.H.
J. Biol. Chem. 267, 12552-12558, 1992
A;Title: Identification and expression of five mutations in the human acid sphingo
opathic and non-neuronopathic forms.
A;Reference number: A42801; MUID:92316934
A;Contents: annotation; characterization of mutations
A;Note: substitution of Ile for 382-Met or Ser for 383-Asn result in complete inac
tion and lead to milder (type B) disease
C;Comment: Two isoforms, neutral and acidic, have been identified. The acidic iso-
C;Comment: For the inactive splice form 2, see PIR:S06958.
C;Genetics:
A;Gene: GDB:SMPD1
A;Cross-references: GDB:128144; OMIM:257200
A;Map position: 11p15.4-11p15.4
A;Introns: 104/3; 362/2; 419/3; 445/2; 494/1
A;Note: a defect in this gene may result in Niemann-Pick disease
C;Function:
A;Description: catalyzes the hydrolysis of sphingomyelin to form phosphorylcholine
C;Superfamily: acid sphingomyelinase; phosphoesterase core homology; saposin repe
C;Keywords: acetylated amino end; alternative splicing; glycoprotein; lysosome; m
F;1-48/Domain: signal sequence #status predicted <SIG>
F;49-629/Product: sphingomyelin phosphodiesterase #status predicted <MAT>
F;81-175/Domain: saposin repeat homology <SAP>
F;200-320/Domain: phosphoesterase core homology <PEC>
F;49/Modified site: acetylated amino end (Asp) (in mature form) #status predicted
F;86,175,335,395,503,520/Binding site: carbohydrate (Asn) (covalent) #status pred.
F;89-165,92-157,120-131/Disulfide bonds: #status predicted
F;251,319,439/Active site: Asp, His, Tyr #status predicted


Query Match                45.8%;    Score 44;  DB 1;  Length 629;
Best Local Similarity      71.4%;    Pred. No. 75;
Matches      10;  Conservative      2;  Mismatches      2;  Indels      0;  Gaps      0;

Qy      3  VLVLALAVALAVGD 16
           || |||:|||:| |
Db     36  VLALALALALSD 49


RESULT    9
S54396
protein L precursor - Peptostreptococcus magnus (strain 3316)
C;Species: Peptostreptococcus magnus
A;Variety: strain 3316
C;Date: 27-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C;Accession: S54396
R;Murphy, J.P.; Duggleby, C.J.; Atkinson, M.A.; Trowern, A.R.; Atkinson, T.; Gowa
Mol. Microbiol. 12, 911-920, 1994
A;Title: The functional units of a peptostreptococcal protein L.
A;Reference number: S54396; MUID:95020613
A;Accession: S54396
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-992 <MUR>
A;Cross-references: EMBL:L04466; NID:g150673; PIDN:AAA67503.1; PID:g150674


Query Match                45.8%;    Score 44;  DB 2;  Length 992;
```

Best Local Similarity   50.0%;   Pred. No. 1.1e+02;
Matches   9;   Conservative   5;   Mismatches   4;   Indels   0;   Gaps   0;

Qy   2 RVLVLALAVALAVGDGSN 19
         ::|: |||| || |:|
Db   6 KLLMAALAGAIVVGGGAN 23

RESULT 10
B75353
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: B75353
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <WHI>
A;Cross-references: GB:AE002021; GB:AE000513; NID:g6459573; PIDN:AAF11348.1; PID:g645956
C;Genetics:
A;Gene: DR1788
A;Map position: 1
C;Superfamily: Deinococcus radiodurans hypothetical protein DR1788

Query Match   44.8%;   Score 43;   DB 2;   Length 127;
Best Local Similarity   69.2%;   Pred. No. 28;
Matches   9;   Conservative   3;   Mismatches   3;   Indels   1;   Gaps   0;

Qy   1 MRVLVLALAVALA 13
       ||:|:|||| |||
Db   1 MRILLVLALAAALA 13

RESULT 11
JC1384
beta-casein precursor - goat
C;Species: Capra aegagrus hircus (domestic goat)
C;Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 23-Feb-1997
R;Roberts, B.; DiTullio, P.; Vitale, J.; Hehir, K.; Gordon, K.
Gene 121, 255-262, 1992
A;Title: Cloning of the goat beta-casein-encoding gene and expression in transgenic mice
A;Reference number: JC1384; MUID:9307039
A;Accession: JC1384
A;Molecule type: DNA
A;Residues: 1-222 <ROB>
A;Cross-references: GB:M90556
C;Genetics:
A;Gene: CSN2
A;Introns: 17/3; 26/3; 35/3; 45/3; 57/3; 221/3
C;Superfamily: beta-casein
C;Keywords: milk; phosphoprotein

Query Match   44.8%;   Score 43;   DB 2;   Length 222;
Best Local Similarity   64.3%;   Pred. No. 44;
Matches   9;   Conservative   3;   Mismatches   3;   Indels   2;   Gaps   0;

Qy   1 MRVLVLALAVALAV 14
        |:||:|| ||||:
Db   1 MKVLILACLVLAI 14

RESULT 12
C70551

hypothetical protein Rv2507 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Accession: C70551
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordo
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyc
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.;
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete ge
A;Reference number: A70500; MUID:98295987
A;Accession: C70551
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-273 <COL>
A;Cross-references: GB:295556; GB:AL123456; NID:g3261778; PIDN:CAB08925.1; PID:e316
C;Genetics:
A;Gene: Rv2507

Query Match   44.8%;   Score 43;   DB 2;   Length 273;
Best Local Similarity   50.0%;   Pred. No. 52;
Matches   8;   Conservative   5;   Mismatches   3;   Indels   0;   Gaps   0;

Qy   3 VLVLALAVALAVGDGS 18
        :||:||| || :||
Db 120 LLVVALVIALVIANGS 135

RESULT 13
H82637
glycerol-3-phosphate dehydrogenase XF1802 [imported] - Xylella fastidiosa (strain 9a
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: H82637
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Se
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: H82637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-346 <SIM>
A;Cross-references: GB:AE004001; GB:AE003849; NID:g9106864; PIDN:AAF84610.1; GSPDB:G
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga,
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carra
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.;
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.;
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Marti
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sav
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki,
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Sil
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1802
C;Superfamily: glycerol-3-phosphate dehydrogenase

Query Match   44.8%;   Score 43;   DB 2;   Length 346;
Best Local Similarity   64.3%;   Pred. No. 63;
Matches   9;   Conservative   1;   Mismatches   1;   Indels   4;   Gaps   0;

Qy   8 LAVALAVGDGSNLG 21
       |||: ||||:||
Db 197 LAVAIGVADGMQLG 210

RESULT 13

```
RESULT  14
T44816
brp protein [imported] - Halobacterium salinarum
C;Species: Halobacterium salinarum
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 31-Jan-2000
C;Accession: T44816
R;Betlach, M.; Friedman, J.; Boyer, H.W.; Pfeifer, F.
Nucleic Acids Res. 12, 7949-7959, 1984
A;Title: Characterization of a halobacterial gene affecting bacterio-opsin gene expressi
A;Reference number: Z22850; MUID:85037955
A;Accession: T44816
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-359 <BET>
A;Cross-references: EMBL:X01081; PIDN:CAA25558.1
A;Note: the source is designated as Halobacterium halobium
C;Genetics:
A;Gene: brp

Query Match     44.8%;  Score 43;  DB 2;  Length 359;
Best Local Similarity   58.8%;  Pred. No. 65;
Matches   10;  Conservative    3;  Mismatches    3;  Indels    4;  Gaps    0;

Qy     2 RVLVLALAVALAVGDGS 18
           |: :||||  |  ||| |:
Db    26 RLSLALAAAFAVGYGA 42


RESULT  15
H75561
ABC transporter, ATP-binding protein, MsbA family - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: H75561
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 , M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: H75561
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-583 <WHI>
A;Cross-references: GB:AE001872; GB:AE001513; NID:g6457750; PIDN:AAF09688.1; PID:g645776
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0096
A;Map position: 1
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match     44.8%;  Score 43;  DB 2;  Length 583;
Best Local Similarity   55.0%;  Pred. No. 98;
Matches   11;  Conservative   1;  Mismatches    1;  Indels    8;  Gaps    0;

Qy     1 MRVLVLALAVALAVGDGSNL 20
           :| | |||  || || ||
Db    18 LRGAVAALAALLLVGTGLNL 37


Search completed: May  3, 2001, 15:53:11
Job time: 90 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        May  5, 2001, 21:25:01 ; Search time 65.18 Seconds
                                        (without alignments)
                                        214.302 Million cell updates/sec

Title:         US-09-426-776-11
Perfect score: 80
Sequence:      1 attcacatccaccagccatg.........gggaccagtccaacttgggg 80

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters:      605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database  :   Issued_Patents_NA:*
              1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
              2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
              3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
              4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
              5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
              6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

              Pred. No. is the number of results predicted by chance to have a
              score greater than or equal to the score of the result being printed,
              and is derived by analysis of the total score distribution.

                                SUMMARIES
```

Result No.	Score	% Query Match	Match	Length	DB	ID	Description
1	28.4	35.5	35.5	10807	1	US-08-206-176-7	Sequence 7, Appli
2	28.4	35.5	35.5	10807	2	US-08-756-506-5	Sequence 5, Appli
3	27.6	34.5	34.5	824	1	US-08-154-019-29	Sequence 29, Appl
4	27.6	34.5	34.5	824	1	US-08-461-333-29	Sequence 29, Appl
5	27.6	34.5	34.5	824	3	US-08-464-167-29	Sequence 29, Appl
6	27.6	34.5	34.5	824	4	US-09-158-313-29	Sequence 29, Appl
7	27.6	34.5	34.5	824	4	US-08-476-798-29	Sequence 29, Appl
c 8	25.4	31.8	31.8	1607	3	US-08-753-007A-3	Sequence 3, Appli
c 9	25.4	31.8	31.8	1607	4	US-09-398-496-3	Sequence 3, Appli
c 10	25.4	31.8	31.8	2467	3	US-08-753-007A-1	Sequence 1, Appli
c 11	25.4	31.8	31.8	2467	4	US-09-398-496-1	Sequence 1, Appli
c 12	24.2	30.2	30.2	2296	2	US-08-899-336-2	Sequence 2, Appli
13	23.8	29.8	29.8	2018	2	US-08-557-973-1	Sequence 1, Appli
14	23.8	29.8	29.8	35081	2	US-08-752-760A-1	Sequence 1, Appli
c 15	23.6	29.5	29.5	274	1	US-08-444-818-63	Sequence 63, Appli
c 16	23.6	29.5	29.5	1788	2	US-08-722-806A-1	Sequence 1, Appli
c 17	23.6	29.5	29.5	3495	1	US-08-446-038B-2	Sequence 2, Appli
18	23.6	29.5	29.5	3495	1	US-08-446-010B-2	Sequence 2, Appli
19	23.6	29.5	29.5	3495	1	US-08-805-445-2	Sequence 2, Appli
20	23.6	29.5	29.5	3495	2	US-08-064-067D-2	Sequence 2, Appli
21	23.6	29.5	29.5	3495	2	US-09-066-208-2	Sequence 2, Appli
22	23.6	29.5	29.5	3629	1	US-08-097-997A-8	Sequence 8, Appli
23	23.6	29.5	29.5	3629	4	US-08-665-574C-8	Sequence 8, Appli
24	23.6	29.5	29.5	3629	4	US-08-946-994-8	Sequence 8, Appli
c 25	23.6	29.5	29.5	6785	4	US-08-444-818-65	Sequence 65, Appl
c 26	23.6	29.5	29.5	7310	4	US-08-444-818-74	Sequence 74, Appl
c 27	23.6	29.5	29.5	8316	4	US-08-444-818-88	Sequence 88, Appl

c 28	23.6	29.5	8987	4	US-08-444-818-137	Sequence 137, App
c 29	23.6	29.5	9185	4	US-08-444-818-122	Sequence 122, App
30	23.6	29.5	9185	4	US-08-444-818-123	Sequence 123, App
c 31	23.6	29.5	9379	4	US-08-444-818-176	Sequence 176, App
c 32	23.6	29.5	9401	1	US-07-910-760-9	Sequence 9, Appli
c 33	23.6	29.5	9401	5	US-08-440-519-9	Sequence 9, Appli
c 34	23.6	29.5	9401	5	PCT-US91-02225-9	Sequence 10, Appli
35		23	3311	2	US-08-239-276-10	Sequence 10, Appli
36		23	3311	2	US-08-468-579B-10	Sequence 10, Appli
37		23	3311	2	US-08-468-577B-10	Sequence 65, Appli
38	22.8	28.5	2458	3	US-08-756-806A-65	Sequence 65, Appli
39	22.8	28.5	2458	3	US-09-143-214-65	Sequence 1, Appli
c 40	22.8	28.5	2793	3	US-08-347-563A-1	Sequence 1, Appli
c 41	22.8	28.5	2793	3	US-08-485-942A-1	Sequence 1, Appli
c 42	22.8	28.5	2793	4	US-08-488-214A-1	Sequence 1, Appli
43	22.8	28.5	2793	4	US-08-488-208A-1	Sequence 1, Appli
44	22.8	28.5	12912	2	US-08-460-751-1	Sequence 1, Appli
45	22.8	28.5	14060	3	US-08-658-136-4	Sequence 4, Appli

```
                                ALIGNMENTS

RESULT 1
US-08-206-176-7
; Sequence 7, Application US/08206176
; Patent No. 5639940
; GENERAL INFORMATION:
     APPLICANT: Garner, Ian
     APPLICANT: Dalrymple, Michael A
     APPLICANT: Prunkard, Donna E
     APPLICANT: Foster, Donald C
     TITLE OF INVENTION: Production of Fibrinogen in Transgenic
     TITLE OF INVENTION: Animals
     NUMBER OF SEQUENCES: 27
     CORRESPONDENCE ADDRESS:
     ADDRESSEE: ZymoGenetics, Inc.
     STREET:  4225 Roosevelt Way, N.E.
     CITY: Seattle
     STATE: WA
     COUNTRY: USA
     ZIP: 98105
     COMPUTER READABLE FORM:
     MEDIUM TYPE: Floppy disk
     COMPUTER: IBM PC compatible
     OPERATING SYSTEM: PC-DOS/MS-DOS
     SOFTWARE: PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
     APPLICATION NUMBER: US/08/206,176
     FILING DATE:
     CLASSIFICATION: 435
     ATTORNEY/AGENT INFORMATION:
     NAME: Parker, Gary E
     REGISTRATION NUMBER: 31-648
     REFERENCE/DOCKET NUMBER: 93-15
     TELECOMMUNICATION INFORMATION:
     TELEPHONE: 206-547-8080 ext 322
     TELEFAX: 206-548-2329
     INFORMATION FOR SEQ ID NO:  7:
     SEQUENCE CHARACTERISTICS:
     LENGTH:  10807 base pairs
     TYPE: nucleic acid
     STRANDEDNESS: double
     TOPOLOGY: linear
     IMMEDIATE SOURCE:
     CLONE: ovine beta-lactoglobulin
US-08-206-176-7

Query Match             35.5%;  Score 28.4; DB 1; Length 10807;
Best Local Similarity   62.9%;  Pred. No. 0.65;
Matches   44;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;
```

```
Qy    9 ccaccagccatgagggtgcttgtactagctcttgctgtggctctcgcagtggggaccag 68
         || |||||||||| |  ||  || || ||   |||| |||| |||||  |||||  | ||||
Db 4249 CCTGCAGCCATGAAGTGCCTCCTCCTTGCCCTGGGCCTGGCCCCTCGCCGTCGGCGTCCAG 4308

Qy   69 tccaacttgg 78
         |||| ||||
Db 4309 GCCATCATCG 4318
```

RESULT 2
US-08-756-506-5
; Sequence 5, Application US/08756506
; Patent No. 5905185
; GENERAL INFORMATION:
;   APPLICANT: Garner, Ian
;   APPLICANT: Cottingham, Ian R.
;   APPLICANT: Temperley, Simon M.
;   APPLICANT: Foster, Donald C.
;   APPLICANT: Sprecher, Cindy A.
;   APPLICANT: Prunkard, Donna E.
;   TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC
;   TITLE OF INVENTION: ANIMALS
;   NUMBER OF SEQUENCES: 25
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: ZymoGenetics, Inc.
;   STREET: 1201 Eastlake Avenue East
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98102
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/756,506
;   FILING DATE:
;   CLASSIFICATION: 800
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Sawislak, Deborah A
;   REGISTRATION NUMBER: 37,438
;   REFERENCE/DOCKET NUMBER: 95-28
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-442-6672
;   TELEFAX: 206-442-6678
;   INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 10807 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-756-506-5

Query Match   35.5%; Score 28.4; DB 2; Length 10807;
Best Local Similarity 62.9%; Pred. No. 0.65;
Matches   44; Conservative   0; Mismatches 26; Indels 0; Gaps 0;

```
Qy    9 ccaccagccatgagggtgcttgtactagctcttgctgtggctctcgcagtggggaccag 68
         || |||||||||| |  ||  || || ||   |||| |||| |||||  |||||  | ||||
Db 4249 CCTGCAGCCATGAAGTGCCTCCTCCTTGCCCTGGGCCTGGCCCCTCGCCGTCGGCGTCCAG 4308

Qy   69 tccaacttgg 78
         |||| ||||
Db 4309 GCCATCATCG 4318
```

RESULT 3
US-08-154-019-29
; Sequence 29, Application US/08154019
; Patent No. 5633076
; GENERAL INFORMATION:

;   APPLICANT: Deboer, Herman A.
;   APPLICANT: Strijker, Rein
;   APPLICANT: Heyneker, Herbert L.
;   APPLICANT: Platenburg, Gerald
;   APPLICANT: Lee, Sang He
;   APPLICANT: Pieper, Frank
;   APPLICANT: Krimpenfort, Paul J.A.
;   TITLE OF INVENTION: Production of Recombinant Polypeptides
;   TITLE OF INVENTION: by Bovine Species and Transgenic Methods
;   NUMBER OF SEQUENCES: 38
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Townsend and Townsend and Crew
;   STREET: One Market Plaza, Steuart Tower, Suite 2000
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94105
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/154,019
;   FILING DATE: 16-NOV-1993
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/461,333
;   FILING DATE: 05-JUN-1995
;   APPLICATION NUMBER: US 08/077,788
;   FILING DATE: 15-JUN-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/895,956
;   FILING DATE: 15-JUN-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/619,131
;   FILING DATE: 27-NOV-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/444,745
;   FILING DATE: 01-DEC-1989
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Liebescheutz, Joe O.
;   REGISTRATION NUMBER: 37,505
;   REFERENCE/DOCKET NUMBER: 16994-003123
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-543-9600
;   TELEFAX: 415-543-5043
;   INFORMATION FOR SEQ ID NO: 29:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 824 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-154-019-29

Query Match   34.5%; Score 27.6; DB 1; Length 824;
Best Local Similarity 63.6%; Pred. No. 0.67;
Matches   42; Conservative   0; Mismatches 24; Indels 0; Gaps 0;

```
Qy   13 cagccatgagggtgcttgtactagctcttgctgtggctctcgcagtggggaccagtcca 72
         |||||||| | |||   |  || |   |||  ||||| |||  |  || |||||  ||| ||
Db  754 CAGCCATGAAGTGCCTCCTCCTTGCCCTGGGCCTGGCCCCTCGCCGTCGGCGTCCAGGCCA 813

Qy   73 acttgg 78
          | | |
Db  814 TCATCG 819
```

RESULT 4
US-08-461-333-29
; Sequence 29, Application US/08461333

Patent No. 5741957
GENERAL INFORMATION:
- ; APPLICANT: Deboer, Herman A.
- ; APPLICANT: Strijker, Rein
- ; APPLICANT: Heyneker, Herbert L.
- ; APPLICANT: Platenburg, Gerald
- ; APPLICANT: Lee, Sang He
- ; APPLICANT: Pieper, Frank
- ; APPLICANT: Krimpenfort, Paul J.A.
- ; TITLE OF INVENTION: Production of Recombinant Polypeptides
- ; TITLE OF INVENTION: by Bovine Species and Transgenic Methods
- ; NUMBER OF SEQUENCES: 38
- ; CORRESPONDENCE ADDRESS:
- ; ADDRESSEE: Townsend and Townsend and Crew
- ; STREET: One Market Plaza, Steuart Tower, Suite 2000
- ; CITY: San Francisco
- ; STATE: California
- ; COUNTRY: USA
- ; ZIP: 94105
- ; COMPUTER READABLE FORM:
- ; MEDIUM TYPE: Floppy disk
- ; COMPUTER: IBM PC compatible
- ; OPERATING SYSTEM: PC-DOS/MS-DOS
- ; SOFTWARE: PatentIn Release #1.0, Version #1.25
- ; CURRENT APPLICATION DATA:
- ; APPLICATION NUMBER: US/08/461,333
- ; FILING DATE: 05-JUN-1995
- ; CLASSIFICATION: 800
- ; PRIOR APPLICATION DATA:
- ; APPLICATION NUMBER: US 08/077,788
- ; FILING DATE: 15-JUN-1993
- ; PRIOR APPLICATION DATA:
- ; APPLICATION NUMBER: US 07/895,956
- ; FILING DATE: 15-JUN-1992
- ; PRIOR APPLICATION DATA:
- ; APPLICATION NUMBER: US 07/619,131
- ; FILING DATE: 27-NOV-1990
- ; PRIOR APPLICATION DATA:
- ; APPLICATION NUMBER: US 07/444,745
- ; FILING DATE: 01-DEC-1989
- ; ATTORNEY/AGENT INFORMATION:
- ; NAME: Liebescheutz, Joe O.
- ; REGISTRATION NUMBER: 37,505
- ; REFERENCE/DOCKET NUMBER: 16994-003123
- ; TELECOMMUNICATION INFORMATION:
- ; TELEPHONE: 415-543-9600
- ; TELEFAX: 415-543-5043
- ; INFORMATION FOR SEQ ID NO: 29:
- ; SEQUENCE CHARACTERISTICS:
- ; LENGTH: 824 base pairs
- ; TYPE: nucleic acid
- ; STRANDEDNESS: single
- ; TOPOLOGY: linear
- ; MOLECULE TYPE: DNA (genomic)

US-08-461-333-29

Query Match 34.5%; Score 27.6; DB 1; Length 824;
Best Local Similarity 63.6%; Pred. No. 0.67;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy   13 cagccatgagggtgcttgctactagctctgtggctctcgcagtggggaccagtcca 72
            |||||||| | || | || ||| | |||| ||||| || ||||| |||
Db  754 CAGCCATGAAGTGCCTCCTGCCTGGGCGTCCGCCCTGTGGCCGTCCAGGCCA 813

Qy   73 acttgg 78
            | | |
Db  814 TCATCG 819
```

RESULT 5
US-08-464-167-29
; Sequence 29, Application US/08464167

Patent No. 6013857
GENERAL INFORMATION:
- ; APPLICANT: Deboer, Herman A.
- ; APPLICANT: Strijker, Rein
- ; APPLICANT: Heyneker, Herbert L.
- ; APPLICANT: Platenburg, Gerald
- ; APPLICANT: Lee, Sang He
- ; APPLICANT: Pieper, Frank
- ; APPLICANT: Krimpenfort, Paul J.A.
- ; TITLE OF INVENTION: Production of Recombinant Polypeptides
- ; TITLE OF INVENTION: by Bovine Species and Transgenic Methods
- ; NUMBER OF SEQUENCES: 38
- ; CORRESPONDENCE ADDRESS:
- ; ADDRESSEE: Townsend and Townsend and Crew
- ; STREET: One Market Plaza, Steuart Tower, Suite 2000
- ; CITY: San Francisco
- ; STATE: California
- ; COUNTRY: USA
- ; ZIP: 94105
- ; COMPUTER READABLE FORM:
- ; MEDIUM TYPE: Floppy disk
- ; COMPUTER: IBM PC compatible
- ; OPERATING SYSTEM: PC-DOS/MS-DOS
- ; SOFTWARE: PatentIn Release #1.0, Version #1.25
- ; CURRENT APPLICATION DATA:
- ; APPLICATION NUMBER: US/08/464,167
- ; FILING DATE: 05-JUN-1995
- ; CLASSIFICATION: 800
- ; PRIOR APPLICATION DATA:
- ; APPLICATION NUMBER: US 08/077,788
- ; FILING DATE: 15-JUN-1993
- ; PRIOR APPLICATION DATA:
- ; APPLICATION NUMBER: US 07/895,956
- ; FILING DATE: 15-JUN-1992
- ; PRIOR APPLICATION DATA:
- ; APPLICATION NUMBER: US 07/619,131
- ; FILING DATE: 27-NOV-1990
- ; PRIOR APPLICATION DATA:
- ; APPLICATION NUMBER: US 07/444,745
- ; FILING DATE: 01-DEC-1989
- ; ATTORNEY/AGENT INFORMATION:
- ; NAME: Liebescheutz, Joe O.
- ; REGISTRATION NUMBER: 37,505
- ; REFERENCE/DOCKET NUMBER: 16994-003124
- ; TELECOMMUNICATION INFORMATION:
- ; TELEPHONE: 415-543-9600
- ; TELEFAX: 415-543-5043
- ; INFORMATION FOR SEQ ID NO: 29:
- ; SEQUENCE CHARACTERISTICS:
- ; LENGTH: 824 base pairs
- ; TYPE: nucleic acid
- ; STRANDEDNESS: single
- ; TOPOLOGY: linear
- ; MOLECULE TYPE: DNA (genomic)

US-08-464-167-29

Query Match 34.5%; Score 27.6; DB 3; Length 824;
Best Local Similarity 63.6%; Pred. No. 0.67;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy   13 cagccatgagggtgcttgctactagctctgtggctctcgcagtggggaccagtcca 72
            |||||||| | || | || ||| | |||| ||||| || ||||| |||
Db  754 CAGCCATGAAGTGCCTCCTGCCTGGGCGTCCGCCCTGTGGCCGTCCAGGCCA 813

Qy   73 acttgg 78
            | | |
Db  814 TCATCG 819
```

RESULT 6
US-09-158-313-29
; Sequence 29, Application US/09158313

Patent No. 6066725
; GENERAL INFORMATION:
; APPLICANT: Deboer, Herman A.
; APPLICANT: Strijker, Rein
; APPLICANT: Heyneker, Herbert L.
; APPLICANT: Platenburg, Gerald
; APPLICANT: Lee, Sang He
; APPLICANT: Pieper, Frank
; APPLICANT: Krimpenfort, Paul J.A.
; TITLE OF INVENTION: Production of Recombinant Polypeptides
; TITLE OF INVENTION: by Bovine Species and Transgenic Methods
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/158,313
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,798
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US-07/895,956
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/619,131
; FILING DATE: 27-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/444,745
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebescheutz, Joe O.
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 16994-003125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-158-313-29

Query Match 34.5%; Score 27.6; DB 3; Length 824;
Best Local Similarity 63.6%; Pred. No. 0.67;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy   13 cagccatgagggtgcttgtactagctcttgctggctcgtgtggctctcgcagtggggaccagtcca 72
         ||||||||| | || ||| || |  ||| ||||||||| |||| ||| |||||| ||| |||
Db  754 CAGCCATGAAGTGCCTCCTTGCCCTGGCCCTGGCCGTGTGGGCCTGGCGTCCAGGCCA 813

Qy   73 acttgg 78
         | ||
Db  814 TCATCG 819
```

RESULT 7
US-08-476-798-29
; Sequence 29, Application US/08476798

Patent No. 6140552
; GENERAL INFORMATION:
; APPLICANT: Deboer, Herman A.
; APPLICANT: Strijker, Rein
; APPLICANT: Heyneker, Herbert L.
; APPLICANT: Platenburg, Gerald
; APPLICANT: Lee, Sang He
; APPLICANT: Pieper, Frank
; APPLICANT: Krimpenfort, Paul J.A.
; TITLE OF INVENTION: Production of Recombinant Polypeptides
; TITLE OF INVENTION: by Bovine Species and Transgenic Methods
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,798
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/077,788
; FILING DATE: 15-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/895,956
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/619,131
; FILING DATE: 27-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/444,745
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebescheutz, Joe O.
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 16994-003125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-476-798-29

Query Match 34.5%; Score 27.6; DB 4; Length 824;
Best Local Similarity 63.6%; Pred. No. 0.67;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy   13 cagccatgagggtgcttgtactagctcttgctggctcgtgtggctctcgcagtggggaccagtcca 72
         ||||||||| | || ||| || |  ||| ||||||||| |||| ||| |||||| ||| |||
Db  754 CAGCCATGAAGTGCCTCCTTGCCCTGGCCCTGGCCGTGTGGGCCTGGCGTCCAGGCCA 813

Qy   73 acttgg 78
         | ||
Db  814 TCATCG 819
```

RESULT 8
US-08-753-007A-3/c
; Sequence 3, Application US/08753007A

; Patent No. 6074841
; GENERAL INFORMATION:
;   APPLICANT: Gearing, David P.
;   APPLICANT: Busfield, Samantha J.
;   TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;   NUMBER OF SEQUENCES: 33
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: US
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/753,007A
;   FILING DATE: 19-NOV-1996
;   CLASSIFICATION: 536
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/699,591
;   FILING DATE: 19-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Fasse, J. Peter
;   REGISTRATION NUMBER: 32,983
;   REFERENCE/DOCKET NUMBER: 07334/022001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617-542-5070
;   TELEFAX: 617-542-8906
;   TELEX:
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1607 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   FEATURE:
;   NAME/KEY: Coding Sequence
;   LOCATION: 79..621
;   OTHER INFORMATION:
US-08-753-007A-3

Query Match   31.8%; Score 25.4; DB 3; Length 1607;
Best Local Similarity   61.2%; Pred. No. 4.4;
Matches   41; Conservative   0; Mismatches   26; Indels   0; Gaps   0;

Qy   8 tccaccagccatgagggtgcttgtactagctcctggctgtgtggccctcgcagtgggggacca 67
             || ||| |||| ||||||| || | || | || | |||| ||||||| |||||
Db   768 TCTGCCGGAGATGATGATGCATCTGCCTCCTCCTCGTTTTTTGGTCTTGCAGTAGGCGACCA 709

Qy   68 gtccaac 74
             | ||
Db   708 CACAGAC 702

RESULT 9
US-09-398-496-3/c
; Sequence 3, Application US/09398496
; Patent No. 6133423
; GENERAL INFORMATION:
;   APPLICANT: Gearing, David P.
;   APPLICANT: Busfield, Samantha J.
;   TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;   AND USES THEREFOR
;   NUMBER OF SEQUENCES: 33
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.

;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: US
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/398,496
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/753,007
;   FILING DATE: 19-NOV-1996
;   APPLICATION NUMBER: 08/699,591
;   FILING DATE: 19-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Fasse, J. Peter
;   REGISTRATION NUMBER: 32,983
;   REFERENCE/DOCKET NUMBER: 07334/022001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617-542-5070
;   TELEFAX: 617-542-8906
;   TELEX:
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1607 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   FEATURE:
;   NAME/KEY: Coding Sequence
;   LOCATION: 79..621
;   OTHER INFORMATION:
US-09-398-496-3

Query Match   31.8%; Score 25.4; DB 4; Length 1607;
Best Local Similarity   61.2%; Pred. No. 4.4;
Matches   41; Conservative   0; Mismatches   26; Indels   0; Gaps   0;

Qy   8 tccaccagccatgagggtgcttgtactagctcctggctgtgtggccctcgcagtgggggacca 67
             || ||| |||| ||||||| || | || | || | |||| ||||||| |||||
Db   768 TCTGCCGGAGATGATGATGCATCTGCCTCCTCCTCGTTTTTTGGTCTTGCAGTAGGCGACCA 709

Qy   68 gtccaac 74
             | ||
Db   708 CACAGAC 702

RESULT 10
US-08-753-007A-1/c
; Sequence 1, Application US/08753007A
; Patent No. 6074841
; GENERAL INFORMATION:
;   APPLICANT: Gearing, David P.
;   APPLICANT: Busfield, Samantha J.
;   TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;   TITLE OF INVENTION: AND USES THEREFOR
;   NUMBER OF SEQUENCES: 33
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: US
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/753,007A
FILING DATE: 19-NOV-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/699,591
FILING DATE: 19-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07334/022001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2467 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79...1893
OTHER INFORMATION:
US-08-753-007A-1

Query Match   31.8%;   Score 25.4; DB 3; Length 2467;
Best Local Similarity 61.2%; Pred. No. 4.9;
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 67

Qy   8 tccaccagccatgagggtgcttgtactagctcctcgtgtgctctcgcagtggggacca 67
         || || | |||| | ||| ||| | |||| ; | |||| ||||| || |||||
Db   685 TCTGCCGGAGATGATGCATCTGCCTCCTCGTTTTTGGTCTTGCAGTAGGCGACCA 626

Qy   68 gtccaac 74
         | ||
Db   625 CACAGAC 619

RESULT 11
US-09-398-496-1/c
; Sequence 1, Application US/09398496
; Patent No. 6133423
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: US
;   ZIP: 02110-2804
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/398,496
;   FILING DATE:
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/753,007

FILING DATE: 19-NOV-1996
APPLICATION NUMBER: 08/699,591
FILING DATE: 19-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07334/022001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2467 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79...1893
OTHER INFORMATION:
US-09-398-496-1

Query Match   31.8%;   Score 25.4; DB 4; Length 2467;
Best Local Similarity 61.2%; Pred. No. 4.9;
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 67

Qy   8 tccaccagccatgagggtgcttgtactagctcctcgtgtgctctcgcagtggggacca 67
         || || | |||| | ||| ||| | |||| ; | |||| ||||| || |||||
Db   685 TCTGCCGGAGATGATGCATCTGCCTCCTCGTTTTTGGTCTTGCAGTAGGCGACCA 626

Qy   68 gtccaac 74
         | ||
Db   625 CACAGAC 619

RESULT 12
US-08-899-336-2/c
; Sequence 2, Application US/08899336
; Patent No. 5955649
; GENERAL INFORMATION:
; APPLICANT: HIROTA, NAOHIKO
; APPLICANT: KIHARA, MAKOTO
; APPLICANT: KURODA, HISAO
; APPLICANT: ITO, KAZUTOSHI
; TITLE OF INVENTION: GENE EXPRESSION REGULATORY DNA,
; TITLE OF INVENTION: EXPRESSION CASSETTE, EXPRESSION VECTOR AND TRANSGENIC
; TITLE OF INVENTION: PLANT
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;   ADDRESSEE: P.C.
;   STREET: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400
;   CITY: ARLINGTON
;   STATE: VIRGINIA
;   COUNTRY: U.S.A.
;   ZIP: 22202
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/899,336
;   FILING DATE: 23-JUL-1997
;   CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP HEI 8-193433
;   FILING DATE: 23-JUL-1996
; ATTORNEY/AGENT INFORMATION:
;   NAME: OBLON, NORMAN F.

```
Qy      8 tccaccagccatgagggtgcttgtactagctcttgctgtggctctcgcagtgggggacc 66
          |||||       |  ||| || || ||  |||||| || ||||||| || |||| |||||
Db    694 TCCACTTCATGCGCCGGTTCTGGAACCAGATCTTGATCTGGCGCTCGGAGAGGCAGAGC 636
```

RESULT 14
US-08-752-760A-1/c
; Sequence 1, Application US/08752760A
; Patent No. 5877011
; GENERAL INFORMATION:
;   APPLICANT: Armentano, Donna
;   APPLICANT: Gregory, Richard J.
;   APPLICANT: Smith, Alan E.
;   TITLE OF INVENTION: CHIMERIC ADENOVIRAL VECTORS
;   NUMBER OF SEQUENCES: 3
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Baker & Botts, L.L.P.
;   STREET: 30 Rockefeller Plaza
;   CITY: New York
;   STATE: NY
;   COUNTRY: U.S.A.
;   ZIP: 10112
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/752,760A
;   FILING DATE: 20-NOV-1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Seide, Rochelle K
;   REGISTRATION NUMBER: 32,300
;   REFERENCE/DOCKET NUMBER: A31385
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 212-705-5000
;   TELEFAX: 212-705-5020
;   TELEX:
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 35081 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-752-760A-1

Query Match   29.8%; Score 23.8; DB 2; Length 35081;
Best Local Similarity   62.7%; Pred. No. 32;
Matches   37; Conservative   0; Mismatches   22; Indels   0; Gaps   0;

```
Qy     22 gggtgcttgtactagctcttgctgtggctctcgcagtgggggaccagtcaacttgggg 80
          |||| || |||| || |||| ||  ||||  || |||  |||     |||  |||||||
Db  16175 GGGTACATGTTCTCGCCCTCGCCGCTTGATCCCGCGGCGTCTCGGCAGCAGCACCCCTGGGG 16117
```

RESULT 15
US-08-444-818-63/c
; Sequence 63, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
;   APPLICANT: Chien, David Y.
;   APPLICANT: Rutter, William J.
;   TITLE OF INVENTION: NANBV Diagnostics and Vaccines
;   NUMBER OF SEQUENCES: 777
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Chiron Corporation
;   STREET: 4560 Horton Street
;   CITY: Emeryville

;   REGISTRATION NUMBER: 24,618
;   REFERENCE/DOCKET NUMBER: 2589-0061-0
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 413-3000
;   TELEFAX: (703) 413-2220
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2296 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-899-336-2

Query Match   30.2%; Score 24.2; DB 2; Length 2296;
Best Local Similarity   66.0%; Pred. No. 12;
Matches   35; Conservative   0; Mismatches   18; Indels   0; Gaps   0;

```
Qy     27 cttgtactagctcttgtgtggctctcgcagtgggggaccagtccaacttggg 79
          |||| ||||||| |||| ||||  || |||  |||| || ||| || ||||||
Db   1198 CTTGTCCTAGCTGTTGCCCTTGTCCTGGCTGCGTGCGGAGAAGTTGCAATTGGG 1146
```

RESULT 13
US-08-557-973-1/c
; Sequence 1, Application US/08557973
; Patent No. 5837507
; GENERAL INFORMATION:
;   APPLICANT: Largman, Corey
;   APPLICANT: Lawrence, H. Jeffrey
;   APPLICANT: Humphries, R. Keith
;   APPLICANT: Sauvageau, Guy
;   TITLE OF INVENTION: HOX-INDUCED ENHANCEMENT OF IN VIVO
;   TITLE OF INVENTION: AND IN VITRO PROLIFERATIVE CAPACITY AND GENE THERAPEUTIC METHO
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson
;   STREET: 2200 Sand Hill Road, Suite 100
;   CITY: Menlo Park
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94025
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: ASCII
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/557,973
;   FILING DATE:
;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Valeta Gregg
;   REGISTRATION NUMBER: 35,127
;   REFERENCE/DOCKET NUMBER: 06510/039001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 322-5070
;   TELEFAX: (415) 854-0875
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2018 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
US-08-557-973-1

Query Match   29.8%; Score 23.8; DB 2; Length 2018;
Best Local Similarity   62.7%; Pred. No. 16;
Matches   37; Conservative   0; Mismatches   22; Indels   0; Gaps   0;

```
;         STATE:  CA
;         COUNTRY:  USA
;         ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;         MEDIUM TYPE:  Floppy disk
;         COMPUTER:  IBM PC compatible
;         OPERATING SYSTEM:  PC-DOS/MS-DOS
;         SOFTWARE:  Patentin Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;         APPLICATION NUMBER:  US/08/444,818
;         FILING DATE:
;         CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;         APPLICATION NUMBER:  US/08/403,590
;         FILING DATE:  14-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;         NAME:  Harbin, Alisa A.
;         REGISTRATION NUMBER:  33,895
;         REFERENCE/DOCKET NUMBER:  0110.002
;    TELECOMMUNICATION INFORMATION:
;         TELEPHONE:  (508)359-3876
;         TELEFAX:  (508)359-3885
; INFORMATION FOR SEQ ID NO:  63:
;    SEQUENCE CHARACTERISTICS:
;         LENGTH:  274 base pairs
;         TYPE:  nucleic acid
;         STRANDEDNESS:  single
;         TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;         NAME/KEY:  CDS
;         LOCATION:  2..274
US-08-444-818-63
```

```
Query Match           29.5%;  Score 23.6;  DB 4;  Length 274;
Best Local Similarity  61.3%;  Pred. No. 12;
Matches  38;  Conservative  0;  Mismatches  24;  Indels  0;  Gaps  0;

Qy     17 catgagggtgcttgactagctcttgctggctctcgcagtggggaccagtccaactt 76
          |||  ||  |||| | ||||    | ||  |||||| |  || || || || |||| -
Db    248 CAAGTGGTTCTATGGAGTAGCAGGCCCCGTAGATCTCGCAATCGAGGGCCTGTTCAAGCT 189

Qy     77 gg 78
          | |
Db    188 GG 187
```

Search completed: May 5, 2001, 22:04:58
Job time: 2397 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2001, 21:25:56 ; Search time 114.18 Seconds
                                         (without alignments)
                                         409.025 Million cell updates/sec

Title: US-09-426-776-11
Perfect score: 80
Sequence: 1 attccacatccaccagccatg..........gggaccagtccaacttgggg 80

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_0401:*
                1: /SIDS6/gcgdata/geneseq/geneseqn/NA1980.DAT:*
                2: /SIDS6/gcgdata/geneseq/geneseqn/NA1981.DAT:*
                3: /SIDS6/gcgdata/geneseq/geneseqn/NA1982.DAT:*
                4: /SIDS6/gcgdata/geneseq/geneseqn/NA1983.DAT:*
                5: /SIDS6/gcgdata/geneseq/geneseqn/NA1984.DAT:*
                6: /SIDS6/gcgdata/geneseq/geneseqn/NA1985.DAT:*
                7: /SIDS6/gcgdata/geneseq/geneseqn/NA1986.DAT:*
                8: /SIDS6/gcgdata/geneseq/geneseqn/NA1987.DAT:*
                9: /SIDS6/gcgdata/geneseq/geneseqn/NA1988.DAT:*
               10: /SIDS6/gcgdata/geneseq/geneseqn/NA1989.DAT:*
               11: /SIDS6/gcgdata/geneseq/geneseqn/NA1990.DAT:*
               12: /SIDS6/gcgdata/geneseq/geneseqn/NA1991.DAT:*
               13: /SIDS6/gcgdata/geneseq/geneseqn/NA1992.DAT:*
               14: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT:*
               15: /SIDS6/gcgdata/geneseq/geneseqn/NA1994.DAT:*
               16: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:*
               17: /SIDS6/gcgdata/geneseq/geneseqn/NA1996.DAT:*
               18: /SIDS6/gcgdata/geneseq/geneseqn/NA1997.DAT:*
               19: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT:*
               20: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:*
               21: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT:*
               22: /SIDS6/gcgdata/geneseq/geneseqn/NA2001.DAT:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	80	100.0	80	21	A28495	O. aureus vitellog
2	80	100.0	152	21	A28499	Vtgss-CAT fusion p
3	80	100.0	152	21	A28500	Vtgss-CAT fusion p
4	80	100.0	155	21	A28502	Vtgss-EGFP fusion
5	80	100.0	204	21	A28496	Vtgss-CrFCES fusio
6	65.4	81.8	204	21	A28507	Vtgss-beta-lactama
7	28.8	36.0	573	21	C79853	Human secreted pro
8	28.4	35.5	10807	16	T03855	Sheep beta-lactogl
9	28.4	35.5	10807	18	T79725	Ovine beta-lactogl
10	27.6	34.5	222	18	T60633	Human beta-lactogl
11	27.6	34.5	824	21	Z87633	Sheep beta-lactogl

						Description
12		27.6	34.5	824	22	C68328	Sheep beta lactogl
13		27.4	34.2	879	21	F16035	Human prostate can
14	c	26.8	33.5	4094	21	251205	Human hypoxia regu
15		26.6	33.5	4248	21	A28450	Human Seladin-1 cD
16		26.6	33.2	2712	21	C46769	Arabidopsis thalia
17		25.4	31.8	29	21	A28506	Primer BspSfor fo
18	c	25.4	31.8	1607	19	V17813	Mus musculus don-1
19	c	25.4	31.8	2467	19	V17812	Mus musculus don-1
20	c		25	987	18	T89157	Oerskovia xanthine
21	c		25	1177	18	T89155	Oerskovia xanthine
22	c		25	1516	18	T89156	Oerskovia xanthine
23		24.8	31.0	868	21	C46522	Zea mays DNA fragm
24	c	24.8	31.0	1190	13	V20662	Hybrid mini-vitell
25		24.8	31.0	1247	13	Q21979	Hybrid mini-vitell
26		24.8	31.0	1260	13	Q21978	Hybrid mini-vitell
27		24.8	31.0	1290	13	Q21976	Hybrid mini-vitell
28		24.8	31.0	1290	13	Q21977	Hybrid mini-vitell
29		24.4	30.5	3394	13	Q21981	Human Ras signalli
30		24.4	30.5	4109	21	D00315	Human cAMP-GEFI al
31		24.2	30.2	1520	21	F32777	Human secreted pro
32	c	24.2	30.2	2296	19	V20662	Barley D-hordein g
33		23.8	29.8	1182	22	F44982	Murine INTERCEPT 2
34	c	23.8	29.8	1182	22	F45050	Murine secreted pr
35	c	23.8	29.8	1182	22	F45051	Murine secreted pr
36	c	23.8	29.8	1182	22	F45052	Murine secreted pr
37	c	23.8	29.8	1182	22	F45053	Murine secreted pr
38	c	23.8	29.8	1846	22	F44981	Murine INTERCEPT 2
39	c	23.8	29.8	1846	22	F45018	Murine secreted pr
40	c	23.8	29.8	1846	22	F45019	Murine secreted pr
41	c	23.8	29.8	1846	22	F45020	Murine secreted pr
42	c	23.8	29.8	1846	22	F45021	Murine secreted pr
43	c	23.8	29.8	2018	20	V68216	HOXB4 cDNA sequenc
44	c	23.8	29.8	2030	21	287687	Rat heart malonyl
45	c	23.8	29.8	2030	21	287688	Rat liver malonyl

ALIGNMENTS

RESULT 1
A28495
ID   A28495 standard; DNA; 80 BP.
XX
AC   A28495;
XX
DT   29-AUG-2000 (first entry)
XX
DE   O. aureus vitellogenin secretory signal sequence coding sequence.
XX
KW   Vtgss; vitellogenin; secretory signal sequence; gene expression;
KW   oestrogen receptor binding protein; systemic circulation; ss.
XX
OS   Oreochromis aureus.
XX
FH   Key   Location/Qualifiers
FT   sig_peptide   18..80
FT   /*tag=   a
XX
PN   WO20026366-A1.
XX
PD   11-MAY-2000.
XX
PF   29-OCT-1999; 99WO-SG00108.
XX
PR   30-OCT-1998; 98US-0106426.
PR   26-OCT-1999; 99US-0426776.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
PA   (LAMT/) LAM T J.
XX
PI   Ding JL, Tan NS, Ho B;
XX
DR   WPI; 2000-365615/31.

DR   P-PSDB; Y92278.
XX
PT   Isolated nucleic acid for assaying for heterologous gene expression,
PT   detecting presence of compound that binds to estrogen receptor or
PT   producing desired protein from host cell comprises nucleotide sequence
PT   encoding secretory signal sequence
XX
PS   Claim 2; Page 36; 73pp; English.
XX
CC   The present sequence encodes a piscine, Oreochromis aureus, vitellogenin
CC   secretory sequence (Vtgss). This and variants that comprise conservative
CC   replacements that retain the biological activities of directing secretion
CC   of a fusion protein from a cell and cleavage of the secretory signal
CC   sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC   fused to either a reporter protein or a lipopolysaccharide-binding
CC   protein coding sequence. The isolated nucleic acid is useful for assaying
CC   for heterologous gene expression, detecting the presence of a compound
CC   that binds to an oestrogen receptor in a sample or producing a desired
CC   protein from a host cell. It can also be used in a method for obtaining
CC   systemic circulation of a desired protein in a transgenic or chimeric
CC   host organism.
XX
SQ   Sequence 80 BP; 14 A; 22 C; 24 G; 20 T; 0 other;

Query Match   100.0%; Score 80; DB 21; Length 80;
Best Local Similarity   100.0%; Pred. No. 5.7e-18;
Matches   80; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 attcacatccaccagccatgagggtgcttgtactagctccttgctgtggctctcgcagtgg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 attcacatccaccagccatgagggtgcttgtactagctccttgctgtggctctcgcagtgg 60

Qy   61 gggaccagtccaacttgggg 80
              ||||||||||||||||||||
Db   61 gggaccagtccaacttgggg 80

RESULT   2
A28499
ID   A28499 standard; DNA; 152 BP.
XX
AC   A28499;
XX
DT   29-AUG-2000 (first entry)
XX
DE   Vtgss-CAT fusion protein (partial) coding sequence.
XX
KW   Vtgss; vitellogenin; secretory signal sequence; gene expression;
KW   oestrogen receptor binding protein; systemic circulation; CAT; ss.
XX
OS   Chimeric - Oreochromis aureus.
OS   Chimeric - Synthetic.
XX
FH   Key   Location/Qualifiers
FT   CDS   51..152
FT   /*tag= a
FT   /transl_except= (pos:123..125, aa:Gln)
FT   /transl_except= (pos:132..134, aa:Val)
FT   /transl_except= (pos:141..143, aa:Trp)
FT   sig_peptide   51..113
FT   /*tag= b
FT   mat_peptide   117..152
FT   /*tag= c
FT   /partial
FT   /note= "CAT N-terminal"
XX
PN   WO200026366-A1.
XX
PD   11-MAY-2000.
XX
PF   29-OCT-1999; 99WO-SG00108.
XX

PR   30-OCT-1998; 98US-0106426.
PR   26-OCT-1999; 99US-0426776.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
PA   (LAMT/) LAM T J.
XX
PI   Ding JL, Tan NS, Ho B;
XX
DR   WPI: 2000-365615/31.
DR   P-PSDB; Y92780.
XX
PT   Isolated nucleic acid for assaying for heterologous gene expression,
PT   detecting presence of compound that binds to estrogen receptor or
PT   producing desired protein from host cell comprises nucleotide sequence
PT   encoding secretory signal sequence
XX
PS   Example 2; Fig 5B; 73pp; English.
XX
CC   A reporter CAT system that uses the piscine, Oreochromis aureus,
CC   vitellogenin secretory sequence (Vtgss), designated pBSVtgCAT was
CC   constructed. A28499 and A28722 comprise the 5'- and 3'-ends of the
CC   construct insert. Vtgss and variants that comprise conservative
CC   replacements that retain the biological activities of directing secretion
CC   of a fusion protein from a cell and cleavage of the secretory signal
CC   sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC   fused to either a reporter protein or a lipopolysaccharide-binding
CC   protein coding sequence. The isolated nucleic acid is useful for assaying
CC   for heterologous gene expression, detecting the presence of a compound
CC   that binds to an oestrogen receptor in a sample or producing a desired
CC   protein from a host cell. It can also be used in a method for obtaining
CC   systemic circulation of a desired protein in a transgenic or chimeric
CC   host organism.
XX
SQ   Sequence 152 BP; 38 A; 38 C; 39 G; 37 T; 0 other;

Query Match   100.0%; Score 80; DB 21; Length 152;
Best Local Similarity   100.0%; Pred. No. 6.6e-18;
Matches   80; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 attcacatccaccagccatgagggtgcttgtactagctccttgctgtggctctcgcagtgg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   34 attcacatccaccagccatgagggtgcttgtactagctccttgctgtggctctcgcagtgg 93

Qy   61 gggaccagtccaacttgggg 80
              ||||||||||||||||||||
Db   94 gggaccagtccaacttgggg 113

RESULT   3
A28500
ID   A28500 standard; DNA; 152 BP.
XX
AC   A28500;
XX
DT   29-AUG-2000 (first entry)
XX
DE   Vtgss-CAT fusion protein (partial) coding sequence.
XX
KW   Vtgss; vitellogenin; secretory signal sequence; gene expression;
KW   oestrogen receptor binding protein; systemic circulation; CAT; ss.
XX
OS   Chimeric - Oreochromis aureus.
OS   Chimeric - Synthetic.
XX
FH   Key   Location/Qualifiers
FT   CDS   51..152
FT   /*tag= a
FT   /transl_except= (pos:123..125, aa:Gln)
FT   /transl_except= (pos:132..134, aa:Val)
FT   /transl_except= (pos:141..143, aa:Trp)
FT   sig_peptide   51..113
FT   /*tag= b

FT   mat_peptide   117..152
FT   /tag= c
FT   /partial
FT   /note= "CAT N-terminal"
XX
PN   WO20026366-A1.
XX
PD   11-MAY-2000.
XX
PF   29-OCT-1999;   99WO-SG00108.
XX
PR   30-OCT-1998;   98US-0106426.
PR   26-OCT-1999;   99US-0426776.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
PA   (LAMT/) LAM T J.
XX
PI   Ding JL, Tan NS, Ho B;
XX
DR   WPI; 2000-365615/31.
DR   P-PSDB; Y92780.
XX
PT   Isolated nucleic acid for assaying for heterologous gene expression,
PT   detecting presence of compound that binds to estrogen receptor or
PT   producing desired protein from host cell comprises nucleotide sequence
PT   encoding secretory signal sequence
XX
PS   Example 2; Fig 5D; 73pp; English.
XX
CC   A reporter CAT system that uses the piscine, Oreochromis aureus,
CC   vitellogenin secretory sequence (Vtgss), designated psp-VtgCAT was
CC   constructed. A28500 and A28723 comprise the 5'- and 3'-ends of the
CC   construct insert. Vtgss and variants that comprise conservative
CC   replacements that retain the biological activities of directing secretion
CC   of a fusion protein from a cell and cleavage of the secretory signal
CC   sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC   fused to either a reporter protein or a lipopolysaccharide-binding
CC   protein coding sequence. The isolated nucleic acid is useful for assaying
CC   for heterologous gene expression, detecting the presence of a compound
CC   that binds to an oestrogen receptor in a sample or producing a desired
CC   protein from a host cell. It can also be used in a method for obtaining
CC   systemic circulation of a desired protein in a transgenic or chimeric
CC   host organism.
XX
SQ   Sequence 152 BP; 39 A; 38 C; 39 G; 36 T; 0 other;

Query Match   100.0%;   Score 80; DB 21; Length 152;
Best Local Similarity 100.0%;   Pred. No. 6.6e-18;
Matches   80; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 attcacatccaccagccatgagggtgcttgactactcttgctgtggctctcgcagtgg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   34 attcacatccaccagccatgagggtgcttgactactcttgctgtggctctcgcagtgg 93

Qy   61 gggaccagtccaacttgggg 80
          ||||||||||||||||||||
Db   94 gggaccagtccaacttgggg 113

RESULT   4
A28502
ID   A28502 standard; DNA; 155 BP.
XX
AC   A28502;
XX
DT   29-AUG-2000 (first entry)
XX
DE   Vtgss-EGFP fusion protein (partial) coding sequence.
XX
KW   Vtgss; vitellogenin; secretory signal sequence; gene expression;
KW   oestrogen receptor binding protein; systemic circulation; EGFP; ss.
XX

OS   Chimeric - Oreochromis aureus.
OS   Chimeric - Synthetic.
XX
FH   Key   Location/Qualifiers
FT   CDS   42..155
FT   /tag= a
FT   sig_peptide   42..104
FT   /tag= b
FT   /note= "O. aureus Vtgss"
FT   mat_peptide   123..155
FT   /tag= c
FT   /note= "EGFP"
XX
PN   WO20026366-A1.
XX
PD   11-MAY-2000.
XX
PF   29-OCT-1999;   99WO-SG00108.
XX
PR   30-OCT-1998;   98US-0106426.
PR   26-OCT-1999;   99US-0426776.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
PA   (LAMT/) LAM T J.
XX
PI   Ding JL, Tan NS,   Ho B;
XX
DR   WPI; 2000-365615/31.
DR   P-PSDB; Y92782.
XX
PT   Isolated nucleic acid for assaying for heterologous gene expression,
PT   detecting presence of compound that binds to estrogen receptor or
PT   producing desired protein from host cell comprises nucleotide sequence
PT   encoding secretory signal sequence
XX
PS   Example 3; Fig 8B; 73pp; English.
XX
CC   A reporter GFP system that uses the piscine, Oreochromis aureus,
CC   vitellogenin secretory sequence (Vtgss), designated pVtgEGFP was
CC   constructed. Vtgss and variants that comprise conservative
CC   replacements that retain the biological activities of directing secretion
CC   of a fusion protein from a cell and cleavage of the secretory signal
CC   sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC   fused to either a reporter protein or a lipopolysaccharide-binding
CC   protein coding sequence. The isolated nucleic acid is useful for assaying
CC   for heterologous gene expression, detecting the presence of a compound
CC   that binds to an oestrogen receptor in a sample or producing a desired
CC   protein from a host cell. It can also be used in a method for obtaining
CC   systemic circulation of a desired protein in a transgenic or chimeric
CC   host organism.
XX
SQ   Sequence 155 BP; 30 A; 45 C; 49 G; 31 T; 0 other;

Query Match   100.0%;   Score 80;   DB 21; Length 155;
Best Local Similarity 100.0%;   Pred. No. 6.6e-18;
Matches   80; Conservative   0; Mismatches   0; Indels   0; Gaps   0:

Qy   1 attcacatccaccagccatgagggtgcttgtactagctcttgctgtggctctcgcagtgg 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 attcacatccaccaccagccatgagggtgcttgtactagctcttgctgtggctctcgcagtgg 84

Qy   61 gggaccagtccaacttgggg 80
          ||||||||||||||||||||
Db   85 gggaccagtccaacttgggg 104

RESULT   5
A28496
ID   A28496 standard; DNA; 204 BP.
XX
AC   A28496;
XX
XX

```
DT   29-AUG-2000  (first entry)
XX
DE   Vtgss-CrFCES fusion protein coding sequence.
XX
KW   Vtgss; vitellogenin; secretory signal sequence; gene expression;
KW   oestrogen receptor binding protein; systemic circulation; factor C; ss.
XX
OS   Chimeric - Oreochromis aureus.
OS   Chimeric - Carcinoscorpius rotundicauda.
XX
FH   Key            Location/Qualifiers
FT   CDS            52..204
FT                  /tag= a
FT                  /transl_except= (pos:199..201, aa:Tyr)
FT                  /transl_except= (pos:202..204, aa:Phe)
FT                  /partial
FT   sig_peptide    52..114
FT                  /tag= b
FT                  /note= "O. aureus vitellogenin secretory signal sequence"
FT   mat_peptide    115..204
FT                  /tag= c
FT                  /note= "C. rotundicauda Factor C LPS-binding domain"
XX
PN   WO200026366-A1.
XX
PD   11-MAY-2000.
XX
PF   29-OCT-1999;   99WO-SG00108.
XX
PR   30-OCT-1998;   98US-0106426.
PR   26-OCT-1999;   99US-0426776.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
PA   (LAMT/) LAM T J.
XX
PI   Ding JL, Tan NS, Ho B;
XX
DR   WPI; 2000-365615/31.
DR   P-PSDB; Y92779.
XX
PT   Isolated nucleic acid for assaying for heterologous gene expression,
PT   detecting presence of compound that binds to estrogen receptor or
PT   producing desired protein from host cell comprises nucleotide sequence
PT   encoding secretory signal sequence
XX
PS   Example 1; Fig 2A; 73pp; English.
XX
CC   The piscine, Oreochromis aureus, vitellogenin secretory sequence (Vtgss)
CC   was fused upstream to the EcoRI-SalI cDNA fragment encoding the
CC   lipopolysaccharide-binding domain of Carcinoscorpius rotundicauda Factor
CC   C CrFCES for expression and secretion of recombinant ES protein from
CC   Drosophila cells. Vtgss and variants that comprise conservative
CC   replacements that retain the biological activities of directing secretion
CC   of a fusion protein from a cell and cleavage of the secretory signal
CC   sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC   fused to either a reporter protein or a lipopolysaccharide-binding
CC   protein coding sequence. The isolated nucleic acid is useful for assaying
CC   for heterologous gene expression, detecting the presence of a compound
CC   that binds to an oestrogen receptor in a sample or producing a desired
CC   protein from a host cell. It can also be used in a method for obtaining
CC   systemic circulation of a desired protein in a transgenic or chimeric
CC   host organism.
XX
SQ   Sequence 204 BP; 48 A; 45 C; 58 G; 53 T; 0 other;

Query Match        100.0%;    Score 80; DB 21;  Length 204;
Best Local Similarity  100.0%;    Pred. No. 7e-18;
Matches  80;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1 attcacatccacaccagccatgagggtgctgtactagctcttgctgtggctctcgcagtgg 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35 attcacatccacaccagccatgagggtgctgtactagctcttgctgtggctctcgcagtgg 94

Qy  61 gggaccagtccaacttggg 80
         |||||||||||||||||||
Db  95. gggaccagtccaacttgggg 114


RESULT   6
A28507
ID   A28507 standard; DNA; 204 BP.
XX
AC   A28507;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Vtgss-beta-lactamase fusion protein (partial) coding sequence.
XX
KW   Vtgss; vitellogenin; secretory signal sequence; gene expression;
KW   oestrogen receptor binding protein; systemic circulation;
KW   beta-lactamase; ss.
XX
OS   Chimeric - Oreochromis aureus.
OS   Chimeric - Synthetic.
XX
FH   Key            Location/Qualifiers
FT   CDS            52..204
FT                  /tag= a
FT                  /partial
FT   sig_peptide    52..114
FT                  /tag= b
FT                  /note= "Vtgss from O. aureus"
FT   mat_peptide    115..204
FT                  /tag= c
FT                  /note= "beta-lactamase mature protein"
XX
PN   WO200026366-A1.
XX
PD   11-MAY-2000.
XX
PF   29-OCT-1999;   99WO-SG00108.
XX
PR   30-OCT-1998;   98US-0106426.
PR   26-OCT-1999;   99US-0426776.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
PA   (LAMT/) LAM T J.
XX
PI   Ding JL, Tan NS, Ho B;
XX
DR   WPI; 2000-365615/31.
DR   P-PSDB; Y92783.
XX
PT   Isolated nucleic acid for assaying for heterologous gene expression,
PT   detecting presence of compound that binds to estrogen receptor or
PT   producing desired protein from host cell comprises nucleotide sequence
PT   encoding secretory signal sequence
XX
PS   Example 6; Fig 14A; 73pp; English.
XX
CC   A reporter beta-lactamase system that uses the Oreochromis aureus,
CC   vitellogenin secretory sequence (Vtgss), designated pBADVtgblacttkana was
CC   constructed. Vtgss and variants that comprise conservative
CC   replacements that retain the biological activities of directing secretion
CC   of a fusion protein from a cell and cleavage of the secretory signal
CC   sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC   fused to either a reporter protein or a lipopolysaccharide-binding
CC   protein coding sequence. The isolated nucleic acid is useful for assaying
CC   for heterologous gene expression, detecting the presence of a compound
CC   that binds to an oestrogen receptor in a sample or producing a desired
CC   protein from a host cell. It can also be used in a method for obtaining
CC   systemic circulation of a desired protein in a transgenic or chimeric
CC   host organism.
XX
SQ   Sequence 204 BP; 49 A; 42 C; 60 G; 53 T; 0 other;
```

Query Match   81.8%;   Score 65.4; DB 21; Length 204;
Best Local Similarity 98.5%; Pred. No. 5.2e-13;
Matches 66; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   14 agccatgagggtgcttgtactagctcctgctgtggctctcgcagtggggaccagtccaa 73
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   48 aaccatgagggtgcttgtactagctcctgctgtggctctcgcagtggggaccagtccaa 107

Qy   74 cttgggg 80
         |||||||
Db   108 cttgggg 114

RESULT 7
C79853
ID C79853 standard; cDNA; 573 BP.
XX
AC C79853;
XX
DT 09-FEB-2001 (first entry)
XX
DE Human secreted protein encoding cDNA for gene 5.
XX
KW Human; secreted protein; cytostatic; antiarthritic; antiasthmatic;
KW immunosuppressive; antiarteriosclerotic; antiinflammatory; nootropic;
KW neuroprotective; antidiabetic; tranquilliser; vulnerary; antibacterial;
KW antipsoriatic; antiarrhythmic; antirheumatic; cardiant; anti-HIV;
KW autoimmune disorder; allergic condition; cardiovascular disorder;
KW cancer; neurological disease; tissue repair; ss.
XX
OS Homo sapiens.
XX
XX WO200055176-A2.
PN WO200055176-A2.
PD 21-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06057.
XX
PR 12-MAR-1999; 99US-0124142.
PR 11-JUN-1999; 99US-0138597.
PR 03-DEC-1999; 99US-0168666.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-638176/61.
DR P-PSDB; B44834.
XX
PT Novel 49 human secreted proteins useful for diagnosis, prevention and
PT treatment of disorders including neurological, cell proliferative,
PT cardiovascular, and autoimmune/inflammatory disorders and microbial
PT infections -
XX
PS Claim 1a; Page 338-339; 405pp; English.
XX
CC This invention describes a novel isolated polypeptide (I) comprising an
CC amino acid sequence at least 95 % identical to a polypeptide sequence
CC selected from 49 polypeptides encoded by polynucleotide sequences
CC included in American Type Culture Collection (ATCC) deposit number
CC 203917, defined in the specification. The products of the invention have
CC cytostatic, antiarthritic, antiasthmatic, immunosuppressive, nootropic,
CC antiarteriosclerotic, antiinflammatory, neuroprotective, antidiabetic,
CC tranquilliser, vulnerary, antibacterial, antipsoriatic, antiarrhythmic,
CC antirheumatic, cardiant and anti-HIV activity. (I) or a nucleic acid (II)
CC encoding (I) is useful for preventing, treating or ameliorating a medical
CC condition and for diagnosing a pathological condition or susceptibility
CC to the condition. (I) is useful for identifying a binding partner which
CC affects the activity of the polypeptide and for identifying an activity
CC in a biological sample. (I), (II) or an antibody (IV) specific to (I) is
CC also useful for treating or preventing a disease, disorder or condition

CC associated with aberrant expression of (I). Diseases treated or diagnosed
CC include immune disorders such as autoimmune diseases, blood protein
CC disorders, anemia, allergic reactions and conditions such as asthma,
CC organ rejection or graft-versus-host disease, inflammation, hyper
CC proliferative disorders, cardiovascular disorders such as arterioarterial
CC fistula, arrhythmias, arteriosclerosis, coronary thrombosis, organ
CC regeneration, cancer, neovascular glaucoma, diabetic retinopathy,
CC rheumatoid arthritis, psoriasis, diseases associated with increased
CC apoptosis that include acquired immunodeficiency syndrome (AIDS),
CC neurological diseases such as Parkinson's disease, viral, bacterial,
CC fungal or parasitic diseases. They are also used to repair, replace or
CC protect tissue damage by congenital defects, to treat trauma, in surgery,
CC including cosmetic plastic surgery, to treat fibrosis, reperfusion injury
CC or systemic cytokine damage, to stimulate chondrocyte growth, to prevent
CC skin aging due to sunburn, to change a mammal's mental state or physical
CC state by influencing biorhythms, cardiac rhythms, depression, memory,
CC stress and for accelerating wound healing. (I), (II) and/or their agonist
CC or antagonist are useful as food additives or preservatives to increase
CC or decrease storage capabilities, fat content, lipid, protein,
CC carbohydrate, vitamin, mineral or other nutritional components. (I) is
CC useful for screening therapeutic compounds. (II) is useful in forensic
CC biology for detecting DNA sequences and as diagnostic probes for
CC detecting the presence of specific mRNA in a particular cell type.
XX
SQ Sequence 573 BP; 139 A; 135 C; 166 G; 133 T; 0 other;

Query Match   36.0%;   Score 28.8; DB 21; Length 573;
Best Local Similarity   60.0%;   Pred. No. 1;
Matches 48; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy   1 attcacatccaccagcatgagggtgcttgtactagctccttgctgtggctctcgcagtgg 60
         | |||| ||| ||| || | ||||| || ||||||| || ||||| |||| ||| || ||
Db   346 aatcaaatgcacacgcatggaagctgcttgtccatgccactggttttgtcactggctcct 405

Qy   61 gggaccgtccaacttgggg 80
         || ||| | |||||||||
Db   406 ggaatccgaccgtgttgggg 425

RESULT 8
T03855
ID T03855 standard; DNA; 10807 BP.
XX
AC T03855;
XX
DT 26-OCT-1996 (first entry)
XX
DE Sheep beta-lactoglobulin coding sequence.
XX
KW Sheep beta-lactoglobulin; signal peptide; promoter; fusion protein;
KW human fibrinogen A-alpha chain; B-beta chain; gamma chain;
KW transgenic animal milk; ds.
XX
OS Ovis aries.
XX
PN WO523868-A1.
XX
PD 08-SEP-1995.
XX
PF 01-MAR-1995; 95WO-US02648.
XX
PR 03-MAR-1994; 94US-0206176.
XX
PA (PHAR-) PHARM PROTEINS LTD.
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Dalrymple MA, Foster DC, Garner I, Prunkard DE;
XX
DR WPI; 1995-320582/41.
XX
PT Production of fibrinogen in transgenic mammals - by introducing DNA
PT segments into the germ line of a non-human mammal and collecting

PT milk from female progeny.
XX
PS Disclosure; Page 69-77; 99pp; English.
XX
CC This sequence encodes sheep beta-lactoglobulin and contains a unique
CC EcoRV site at nucleotide 4245 in the 5' untranslated region of the
CC gene. This site allows insertion of additional DNA sequences under
CC the control of the beta-lactoglobulin promoter 3' to the
CC transcription initiation site. The promoter and signal peptide may
CC be used for mamma tissue-specific gene expression and secretion of
CC human fibrinogen A-chain (see T03853), B-beta chain (T03852) and
CC gamma chain (T03854) in non-human transgenic animals. A region of
CC at least the proximal 406 bp of 5' flanking sequences (nucleotides
CC 3844-4257) is generally used, although larger sequences such as
CC nucleotides 1-4257 are preferred. Furthermore, the region
CC surrounding the initiation ATG of one or more of the human
CC fibrinogen sequences may be replaced with the corresponding
CC beta-lactoglobulin sequences providing a putative tissue-specific
CC initiation environment to enhance expression.
XX
SQ Sequence 10807 BP; 2215 A; 3065 C; 2985 G; 2542 T; 0 other;

Query Match   35.5%;   Score 28.4; DB 16;   Length 10807;
Best Local Similarity   62.9%;   Pred. No. 2.6;
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy   9 ccaccagccatgagggtgcttgtactagctcttgctggctctcgcagtggggaccag 68
         || |||||||||| | || ||| || |||| |||| |||| || |||||
Db 4249 cctgcagccatgaagtgcctccctgcttgccctggcctggcctgggtgcctgggtccag 4308

Qy   69 tccaacttgg 78
         ||| |||
Db 4309 gccatcatcg 4318

RESULT 9
T79725
ID   T79725 standard; DNA; 10807 BP.
XX
AC   T79725;
XX
DT   11-DEC-1997 (first entry)
XX
DE   Ovine beta-lactoglobulin gene.
XX
KW   Protein C; transgenic animal; sheep; rabbit; cattle; goat; milk;
KW   blood clotting; anticoagulant; beta-lactoglobulin; promoter; ss.
XX
OS   Ovis aries.
XX
FH   Key   Location/Qualifiers
FT   5'UTR   1..4257
FT   /*tag= a
XX
PN   WO9720043-A1.
XX
PD   05-JUN-1997.
XX
PF   26-NOV-1996; 96WO-US18866.
XX
PR   13-JUN-1996; 96US-0019692.
PR   30-NOV-1995; 95US-0565074.
XX
PA   (PPLT-) PPL THERAPEUTICS.
PA   (ZYMO) ZYMOGENETICS INC.
XX
PI   Cottingham I, Foster DC, Garner I, Prunkard DE;
PI   Sprecher CA, Temperley SM;
XX
DR   WPI; 1997-310599/28.
XX
PT   Production of protein C in transgenic animal - useful for high
PT   quantity protein C production with therapeutic value
XX
PS   Disclosure; Page 66-74; 99pp; English.
XX
CC   This DNA sequence comprises the ovine beta-lactoglobulin gene
CC   including the promoter region. A DNA segment encompassing the
CC   5' flanking promoter region can be utilised in novel constructs
CC   for the expression of human protein C in the milk of a transgenic
CC   animal. A claimed method involves: (a) providing a DNA construct
CC   comprising DNA encoding a secretion signal and a protein C
CC   propeptide, operably linked to DNA encoding two-chain cleavage
CC   site-modified protein C (see W25085-86), the 2 DNA sequences being
CC   linked to elements required for protein C expression in a mammary
CC   gland of a host female animal, such as beta-lactoglobulin gene
CC   promoter; and (b) using the DNA construct to breed a transgenic
CC   animal (esp. sheep, rabbit, cattle, goat) that produces protein C
CC   in its milk, at least 90% of the protein C being in the two-chain
CC   form.
XX
SQ   Sequence 10807 BP; 2214 A; 3066 C; 2985 G; 2542 T; 0 other;

Query Match   35.5%;   Score 28.4; DB 18;   Length 10807;
Best Local Similarity   62.9%;   Pred. No. 2.6;
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy   9 ccaccagccatgagggtgcttgtactagctcttgctggctctcgcagtggggaccag 68
         || |||||||||| | || ||| || |||| |||| |||| || |||||
Db 4249 cctgcagccatgaagtgcctccctgcttgccctggcctggcctgggtgcctgggtccag 4308

Qy   69 tccaacttgg 78
         ||| |||
Db 4309 gccatcatcg 4318

RESULT 10
T60633
ID   T60633 standard; DNA; 222 BP.
XX
AC   T60633;
XX
DT   23-DEC-1997 (first entry)
XX
DE   Human beta-lactoglobulin 5' fragment used to construct hybrid genes.
XX
KW   Human serum albumin; beta-lactoglobulin; hybrid; construct;
KW   transgenic mammal; mammary gland; exogenous DNA; BLG; HSA; milk; ds.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   primer_bind   complement (12..46)
FT   /*tag= a
FT   /label= Oligo_#1
FT   exon   77..212
FT   /*tag= b
FT   /number= 1
FT   primer_bind   113..163
FT   /*tag= c
FT   /label= Oligo_#2
FT   primer_bind   complement (145..194)
FT   /*tag= d
FT   /label= Oligo_#3
FT   intron   213..222
FT   /*tag= e
FT   /number= 1
FT   /note= "Start of intron 1"
XX
PN   EP771874-A2.
XX
PD   07-MAY-1997.
XX
PF   96EP-0117613.

```
XX
PR   03-NOV-1995;   95IL-0115873.
XX
PA   (ISRA ) ISRAEL MIN AGRIC.
XX
PI   Barash I, Hurwitz DR, Nathan M, Shani M;
XX
DR   WPI; 1997-247416/23.
XX
PT   Beta-lactoglobulin hybrid DNA constructs - for producing transgenic
PT   mammals expressing exogenous DNA in mammary glands
XX
PS   Example 1; Fig 1; 16pp; English.
XX
CC   The present sequence represents a fragment of beta-lactoglobulin (BLG)
CC   used in the construction of hybrid genes. The hybrid construct comprises
CC   BLG 5'-flanking sequences and BLG intragenic sequences in conjunction
CC   with a desired cDNA or minigene or gene for targetting expression of the
CC   encoding DNA. The construct is used for producing transgenic mammals
CC   that secrete human serum albumin (HSA) in their milk. By using BLG, the
CC   expression of HSA DNA is targeted to cells in the mammary gland.
XX
SQ   Sequence 222 BP; 37 A; 79 C; 66 G; 40 T; 0 other;

Query Match      34.5%;   Score 27.6;  DB 18;  Length 222;
Best Local Similarity   63.6%;   Pred. No. 2.1;
Matches   42;  Conservative   0;  Mismatches   24;  Indels   0;  Gaps   0;

Qy      13 cagccatgagggtgctgtactagctcttgctggctctcgcagtggggaccagtcca 72
           ||||||||| || ||| ||||| |||||| || | |||| ||| || |||| |||
Db     112 cagccatgaagtgcctcctgcttgccctggcctgcccctcgtggcgtccaggcca 171

Qy      73 acttgg 78
           | | |
Db     172 tcatcg 177


RESULT 11
Z87633
ID   Z87633 standard; DNA; 824 BP.
XX
AC   Z87633;
XX
DT   04-MAY-2000 (first entry)
XX
DE   Sheep beta-lactoglobulin (betaLG) gene fragment.
XX
KW   Transgenic bovine; transgene; milk; serum protein; industrial enzyme;
KW   infant formulation; lactoferrin; intestinal tract infection; lysozyme;
KW   iron absorption; albumin; antibacterial; iron sequestration; betaLG;
KW   beta-lactoglobulin; sheep; ds.
XX
OS   Ovis sp.
XX
PN   US6013857-A.
XX
PD   11-JAN-2000.
XX
PF   05-JUN-1995;   95US-0464167.
XX
PR   16-NOV-1993;   93US-0154019.
PR   01-DEC-1989;   89US-0444745.
PR   27-NOV-1990;   90US-0619131.
PR   15-JUN-1992;   92US-0898956.
PR   15-JUN-1993;   93US-0077788.
XX
PA   (PHAR-) PHARMING BV.
XX
PI   Deboer HA, Heyneker HL, Platenburg G, Krimpenfort PJA, Lee SH;
PI   Pieper F, Strijker R;
XX
DR   WPI; 2000-146563/13.
```

```
XX
PT   Transgenic cattle containing transgene controlled by mammary-specific
PT   regulator, for expressing proteins in the milk, particularly human
PT   lactoferrin for infant feeding formulations -
XX
PS   Example 20; Fig 24; 92pp; English.
XX
CC   The invention provides a transgenic bovine in which the somatic and germ
CC   cells contain a transgene comprising a regulatory sequence from a gene
CC   expressed in mammary glands, DNA encoding a signal sequence and DNA
CC   encoding a naturally occurring heterologous polypeptide. The transgenic
CC   bovine, or its descendants, produce milk containing the heterologous
CC   polypeptide. The transgenic bovines are used to express human milk and
CC   serum proteins or industrial enzymes, specifically for infant
CC   formulations that contain human lactoferrin for control of intestinal
CC   tract infections and to improve iron absorption, particularly when
CC   potentiated by human lysozyme. The polypeptide expressed may also be
CC   human albumin, used as a plasma extender. The polypeptide expressed in
CC   milk of the transgenic bovine requires little if any purification before
CC   human consumption and is expressed at significantly higher levels than in
CC   transgenic mice or sheep. Large polypeptides that are difficult to
CC   express in other systems can also be expressed. The present sequence
CC   represents a sheep beta-lactoglobulin gene fragment.
XX
SQ   Sequence 824 BP; 138 A; 278 C; 245 G; 163 T; 0 other;

Query Match      34.5%;   Score 27.6;  DB 21;  Length 824;
Best Local Similarity   63.6%;   Pred. No. 2.8;
Matches   42;  Conservative   0;  Mismatches   24;  Indels   0;  Gaps   0;

Qy      13 cagccatgagggtgctgtactagctcttgctggctctcgcagtggggaccagtcca 72
           ||||||||| || ||| ||||| |||||| || | |||| ||| || |||| |||
Db     754 cagccatgaagtgcctcctgcttgccctggcctgcccctcgtggcgtccaggcca 813

Qy      73 acttgg 78
           | | |
Db     814 tcatcg 819


RESULT 12
C68328
ID   C68328 standard; DNA; 824 BP.
XX
AC   C68328;
XX
DT   20-FEB-2001 (first entry)
XX
DE   Sheep beta lactoglobulin DNA.
XX
KW   Lactoferrin; mammary; milk; ss.
XX
OS   Ovis aries.
XX
PN   US6140552-A.
XX
PD   31-OCT-2000.
XX
PF   07-JUN-1995;   95US-0476798.
XX
PR   16-NOV-1993;   93US-0154019.
PR   01-DEC-1989;   89US-0444745.
PR   27-NOV-1990;   90US-0619131.
PR   15-JUN-1992;   92US-0898956.
PR   15-JUN-1993;   93US-0077788.
XX
PA   (PHAR-) PHARMING BV.
XX
PI   Strijker R, Heyneker HL, Platenburg G, Pieper F, Krimpenfort PJA;
PI   Lee SH, Deboer HA;
XX
DR   WPI; 2001-040323/05.
XX
```

PT New transgenic bovine whose mammary gland cells contain DNA encoding a
PT signal sequence, and a polypeptide of interest and an expression
PT regulatory sequence, for producing polypeptides in bovine milk -
XX
PS Disclosure; Fig 24; 88pp; English.
XX
CC The present invention relates to a transgenic or chimeric bovine whose
CC mammary gland cells contain a construct encoding a signal sequence, a
CC polypeptide of interest and a regulatory sequence that promotes
CC expression of the DNA sequence. The transgenic or chimeric bovine is
CC useful for producing recombinant polypeptides in milk of female
CC transgenic mammals. The recombinant polypeptide may be used in food
CC formulations, particularly in infant formula having either
CC nutritional or beneficial value. An infant formula containing human
CC lactoferrin from the transgenic bovine milk provides bacteriostatic
CC effect, which aids in controlling diarrhoea in newborn. Recombinant
CC polypeptides may also be used to supplement common diet formulations.
XX
SQ Sequence 824 BP; 138 A; 278 C; 245 G; 163 T; 0 other;

Query Match   34.5%; Score 27.6; DB 22; Length 824;
Best Local Similarity   63.6%; Pred. No. 2.8;
Matches   42; Conservative   0; Mismatches   24; Indels   0; Gaps   0:

Qy   13 cagccatgagggtgcttgtacatagctcttgctgtggctctcgcagtggggaccagtcca 72
          |||||||||||| || ||||||| |||||| |||||| || |||||| ||| || || |||
Db   754 cagccatgaagtgcctcctgcttgccctgggcctggcctggcctggccctgtggcgtccaggcca 813

Qy   73 acttgg 78
          ||| ||
Db   814 tcatcg 819

RESULT 13
F16035/c
ID F16035 standard; cDNA; 879 BP.
XX
AC F16035;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:470.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM.
XX
DR WPI; 2000-587513/55.
DR P-PSDB; B56832.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX

PS Claim 1; Page 943; 2338pp; English.
XX
CC F15566 to F16505 encode the human prostate cancer associated proteins,
CC called prostate cancer antigens, given in B56363 to B57302. The prostate
CC cancer antigens can have neuroprotective, cytostatic, cardioactive,
CC immunomodulatory, muscular, vulnerary, gastrointestinal, nephrotropic,
CC antiinfective, gynaecological and antibacterial activities, and can be
CC used in gene therapy. The prostate cancer antigen polynucleotides may be
CC used for detection of prostate cancer, chromosome identification, as
CC chromosome markers, and for numerous other diagnostic or research
CC purposes. The prostate cancer antigens may be used to treat disorders
CC such as neural, immune, muscular, reproductive, gastrointestinal,
CC pulmonary, cardiovascular, renal, and proliferative disorders, wounds,
CC and infectious diseases. F16506 to F16514 to B57303 represent sequences
CC used in the exemplification of the present invention.
XX
SQ Sequence 879 BP; 172 A; 329 C; 257 G; 116 T; 5 other;

Query Match   34.2%; Score 27.4; DB 21; Length 879;
Best Local Similarity   69.8%; Pred. No. 3.3;
Matches   37; Conservative   0; Mismatches   16; Indels   0; Gaps   0:

Qy   23 ggtgcttgtactagctcttgctgtggctctcgcagtggggaccagtccaact 75
          |||||||| || ||| || |||| || || || |||||||| |||||||| ||
Db   83 GGTCTGGCACCAGGACTGGCACTGCTGCTGGTCCGTGTCCGTGGGGAGCAGTCCATCT 31

RESULT 14
Z51205
ID Z51205 standard; cDNA; 4094 BP.
XX
AC Z51205;
XX
DT 06-JUN-2000 (first entry)
XX
DE Human hypoxia regulatory protein 2-2-83 encoding cDNA.
XX
KW Hypoxia; 2-2-83; anti-hypoxic; vasotropic; neuroprotective; ischaemia;
KW antiapoptotic; angiogenic; tumourigenic; diagnosis; treatment;
KW apoptosis; angiogenesis; neurite outgrowth; tumourigenesis;
KW gene therapy; human; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   37..1587
FT   /*tag= a
FT   /product= "2-2-83 protein"
XX
PN WO200012139-A1.
XX
PD 09-MAR-2000.
XX
PF 27-AUG-1999; 99WO-US20393.
XX
PR 27-AUG-1998; 98US-0098158.
PR 05-MAY-1999; 99US-0132684.
XX
PA (QUAR-) QUARK BIOTECH INC.
PA (KOHN/) KOHN K.
XX
PI Feinstein E, Einat P, Skaliter R;
XX
DR WPI; 2000-237786/20.
DR P-PSDB; Y70123.
XX
PT Polynucleotide sequence comprising a hypoxia response gene, useful for
PT treating and diagnosing hypoxia and ischemia -
XX
PS Claim 7; Fig 2; 80pp; English.
XX
CC The present sequence is the human hypoxia response regulatory cDNA,

Mon May 7 10:36:06 2001

us-09-426-776-11.rng

CC   which encodes 2-2-83 protein. The expression of this gene is modulated
CC   by hypoxic conditions. 2-2-83 protein has anti-hypoxic, vasotropic,
CC   neuroprotective, antiapoptotic, angiogenic and tumourigenic activity.
CC   It can be used for diagnosing and treating hypoxia and ischaemia and
CC   regulating apoptosis, angiogenesis, neurite outgrowth and
CC   tumourigenesis. 2-2-83 nucleotide sequence may be used in gene therapy
CC   and is a useful source of diagnostic probes to identify hypoxia
CC   associated pathologies.
XX
SQ   Sequence 4094 BP; 890 A; 1128 C; 1113 G; 963 T; 0 other;

Query Match   33.5%; Score 26.8; DB 21; Length 4094;
Best Local Similarity   68.5%; Pred. No. 7.1;
Matches   37; Conservative   0; Mismatches   17; Indels   0; Gaps   0;

Qy   2 ttcacatccaccagccatgagggtgcttgctactagctcttgctgtggctctcgc 55
           || |||||||||| || ||| ||| | ||| || | |||||||| |||||||||
Db   113 tgctcatccaccagcgctgggtgttcgtgtgcctcttccctgccgcctctcgc 166

RESULT 15
A28450
ID   A28450 standard; cDNA; 4248 BP.
XX
AC   A28450;
XX
DT   29-AUG-2000 (first entry)
XX
DE   Human Seladin-1 cDNA.
XX
KW   Seladin-1; Alzheimer's disease; Parkinson's disease; Neuroprotective;
KW   nootropic; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   100..1650
FT   /*tag= a
XX
PN   EP1002862-A1.
XX
PD   24-MAY-2000.
XX
PF   12-NOV-1998; 98EP-0121478.
XX
PR   12-NOV-1998; 98EP-0121478.
XX
PA   (NITS/) NITSCH R M.
XX
DR   WPI; 2000-341710/30.
DR   P-PSDB; Y92730.
XX
PT   Novel isolated Seladin-1 polypeptide useful in the diagnosis, prognosis
PT   and treatment of neurological diseases, e.g. Alzheimer's disease and
PT   Amyotrophic lateralsclerosis
XX
PS   Claim 1; Fig 14; 47pp; English.
XX
CC   The present cDNA encodes seladin-1. The cDNA was isolated by
CC   differential display PCR using total RNA from post-mortem brain tissues
CC   from Alzheimer's disease patients and control subjects. The seladin-1
CC   cDNA or protein reduces or prevents the degeneration of neurons and slows
CC   brain amyloid formation. They may be used to diagnose or prognose a
CC   neurological disease or to evaluate a treatment for a neurological
CC   disease (claimed). Neurological diseases treatable by seladin-1 cDNA or
CC   protein include Alzheimer's disease, Parkinson's disease, Huntingdon's
CC   disease, Amyotrophic lateralsclerosis and Pick's disease.
XX
SQ   Sequence 4248 BP; 920 A; 1171 C; 1163 G; 994 T; 0 other;

Query Match   33.5%; Score 26.8; DB 21; Length 4248;

Best Local Similarity   68.5%;   Pred. No. 7.2;
Matches   37; Conservative   0; Mismatches   17; Indels   0; Gaps   0;

Qy   2 ttcacatccaccagccatgagggtgcttgctactagctcttgctgtggctctcgc 55
           | ||||||||||| || ||| || ||| || | || || |||||||||||
Db   -176 tgctcatccaccagcgctgggtgttcgtgtgcctcttccctgccgcctctcgc 229

Search completed: May   5, 2001, 22:07:02
Job time: 2466 sec

```
              GenCore version 4.5
      Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     May  3, 2001, 15:51:41 ; Search time 8.3 Seconds
                                     (without alignments)
                                     86.671 Million cell updates/sec

Title:         US-09-426-776-10
Perfect score: 96
Sequence:      1 MRVLVLALAVALAVGDGSNLG 21

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters:     93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     SwissProt_39:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Match Length	DB	ID	Description
1	55	57.3	674	1	GSPD_VIBCH	P45779 vibrio chol
2	53	55.2	678	1	GSPD_AERHY	P31780 aeromonas h
3	53	55.2	678	1	GSPD_AERSA	P45778 aeromonas s
4	53	55.2	1687	1	VIT2_FUNHE	Q98893 fundulus he
5	45	46.9	945	1	AT1K_HUMAN	060423 homo sapien
6	45	45.8	629	1	ASM_HUMAN	P17405 homo sapien
7	43	44.8	222	1	CASB_CAPHI	P33048 capra hircu
8	43	44.8	262	1	APA1_BRARE	042363 brachydanio
9	43	44.8	683	1	BGH3_HUMAN	Q15582 homo sapien
10	43	44.8	683	1	BGH3_RABIT	Q95215 oryctolagus
11	43	44.8	1148	1	AT1B_MOUSE	P98200 mus musculu
12	43	44.8	1251	1	AT1C_HUMAN	043520 homo sapien
13	42.5	44.3	80	1	YJBE_ECOLI	P32686 escherichia
14	42	43.8	168	1	PBP_MANSE	P18959 manduca sex
15	42	43.8	258	1	APA1_SALSA	P27007 salmo salar
16	42	43.8	378	1	CYOB_HAEIN	P45020 haemophilus
17	42	43.8	550	1	YCHM_ECOLI	P40877 escherichia
18	42	43.8	654	1	GSPD_ECOLI	P45758 escherichia
19	41	42.7	180	1	LACB_CAPHI	P02756 capra hircu
20	41	42.7	180	1	LACB_SHEEP	P02757 ovis aries
21	41	42.7	222	1	CASB_SHEEP	P11839 ovis aries
22	41	42.7	224	1	CASB_BOVIN	P02666 bos taurus
23	41	42.7	224	1	CASB_BUBBU	Q9tsi0 bubalus bub
24	41	42.7	226	1	CASB_HUMAN	P05814 homo sapien
25	41	42.7	228	1	CASB_RABIT	P09116 oryctolagus
26	41	42.7	232	1	CASB_CAMDR	Q9tvd0 camelus dro
27	41	42.7	312	1	HO2_RABIT	P43242 oryctolagus
28	41	42.7	387	1	PAB_PEPMA	Q51911 peptostrept
29	41	42.7	397	1	MML6_MYCTU	Q10773 mycobacteri
30	41	42.7	755	1	COMP_RAT	P35444 rattus norv
31	41	42.7	958	1	MML1_MYCTU	P95211 mycobacteri
32	41	42.7	1122	1	ADP1_MYCGA	Q49379 mycoplasma
33	41	42.7	1163	1	AT5C_HUMAN	060312 homo sapien
34	41	42.7	1508	1	AT5A_MOUSE	054827 mus musculu
35	40.5	42.2	81	1	ATPL_SYNP6	P08445 synechococc
36	40	41.7	160	1	GLB2_CHITH	P02222 chironomus
37	40	41.7	166	1	PAL_PSEPU	P43036 pseudomonas
38	40	41.7	185	1	UPK2_BOVIN	Q08537 bos taurus
39	40	41.7	258	1	IBP4_HUMAN	P22692 homo sapien
40	40	41.7	266	1	PTNC_ECOLI	P08187 escherichia
41	40	41.7	322	1	SERB_ECOLI	P06862 escherichia
42	40	41.7	362	1	ALEU_HORVU	P05167 hordeum vul
43	40	41.7	499	1	PITA_ECOLI	P37308 escherichia
44	40	41.7	499	1	PITB_ECOLI	P43676 escherichia
45	40	41.7	516	1	GUX1_NEUCR	P38676 neurospora

ALIGNMENTS

```
RESULT 1
GSPD_VIBCH
ID   GSPD_VIBCH     STANDARD;     PRT;   674 AA.
AC   P45779;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   GENERAL SECRETION PATHWAY PROTEIN D PRECURSOR (CHOLERA TOXIN SECRETION
DE   PROTEIN EPSD).
GN   EPSD OR VC2733.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR TRH7000;
RA   Overbye L.J.;
RT   "Organization of the general secretion pathway genes in Vibrio
RT   cholerae.";
RL   Thesis (1994), Michigan State University / East Lansing, U.S.A.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T.R., Fleishmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- FUNCTION: REQUIRED FOR SECRETION OF CHOLERA TOXIN THROUGH THE
CC       OUTER MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT (POSSIBLY NATURAL) IN POSITION 55.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L33796; AAA58785.1; -.
DR   EMBL; AE004338; -; NOT_ANNOTATED_CDS.
DR   TIGR; VC2733; -.
DR   InterPro; IPR000016; -.
DR   InterPro; IPR001775; -.
DR   Pfam; PF00263; Bac_GSPproteins; 1.
DR   PRINTS; PR00811; BCTERIALGSPD.
```

```
DR   PROSITE; PS00875; T2SP_D; 1.
KW   Transport; Outer membrane; Signal.
FT   SIGNAL        1     24    POTENTIAL.
FT   CHAIN        25    674    GENERAL SECRETION PATHWAY PROTEIN D.
FT   CONFLICT     89     89    V -> A (IN REF. 1).
FT   CONFLICT    144    144    R -> P (IN REF. 1).
SQ   SEQUENCE  674 AA;  73469 MW;  3D77B891A59E6223 CRC64;

Query Match             57.3%;    Score 55;  DB 1;  Length 674;
Best Local Similarity   60.0%;    Pred. No. 2.1;
Matches   12;  Conservative   3;  Mismatches   5;  Indels   0;  Gaps   0;

Qy     2 RVLVLALAVGDGSNLG 21
         :|| :| | :| ||| |||
Db   343 QVLIEALIVEMAEGDGINLG 362

RESULT   2
GSPD_AERHY
ID   GSPD_AERHY       STANDARD;       PRT;       678 AA.
AC   P31780;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   GENERAL SECRETION PATHWAY PROTEIN D PRECURSOR.
GN   EXED.
OS   Aeromonas hydrophila.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC   Aeromonas.
OX   NCBI_TaxID=644;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AH65;
RA   Howard S.P.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 408-678 FROM N.A.
RC   STRAIN=AH65;
RX   MEDLINE=92349963; PubMed=1640836;
RA   Jiang B., Howard S.P.;
RT   "The Aeromonas hydrophila exeE gene, required both for protein
RT   secretion and normal outer membrane biogenesis, is a member of a
RT   general secretion pathway.";
RL   Mol. Microbiol. 6:1351-1361(1992).
CC   -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC       EXPORT OF PROTEINS.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.
CC   -----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------
DR   EMBL; X66504; CAA47124.1; -.
DR   PIR; S22668; S22668.
DR   InterPro; IPR000016; -.
DR   InterPro; IPR001775; -.
DR   Pfam; PF00263; Bac_GSPproteins; 1.
DR   PRINTS; PR00811; BCTERIALGSPD.
DR   PROSITE; PS00875; T2SP_D; 1.
KW   Transport; Outer membrane; Signal.
FT   SIGNAL        1     25    POTENTIAL.
FT   CHAIN        26    678    GENERAL SECRETION PATHWAY PROTEIN D.
SQ   SEQUENCE  678 AA;  72451 MW;  43B33A28861B0238 CRC64;

Query Match             55.2%;    Score 53;  DB 1;  Length 678;
Best Local Similarity   60.0%;    Pred. No. 4;
```

```
Matches   12;  Conservative   3;  Mismatches   5;  Indels   0;  Gaps   0;

Qy     2 RVLVLALAVGDGSNLG 21
         :||| |: | :| ||| |||
Db   353 QVLVEAIIVEIADGDGLNLG 372

RESULT   3
GSPD_AERSA
ID   GSPD_AERSA       STANDARD;       PRT;       678 AA.
AC   P45778;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   GENERAL SECRETION PATHWAY PROTEIN D PRECURSOR.
GN   EXED.
OS   Aeromonas salmonicida.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC   Aeromonas.
OX   NCBI_TaxID=645;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCMB 1102;
RX   MEDLINE=95309729; PubMed=7789814;
RA   Karlyshev A.V., Macintyre S.;
RT   "Cloning and study of the genetic organization of the exe gene
RT   cluster of Aeromonas salmonicida.";
RL   Gene 158:77-82(1995).
CC   -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC       EXPORT OF PROTEINS.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.
CC   -----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------
DR   EMBL; X80505; CAA56668.1; -.
DR   InterPro; IPR000016; -.
DR   InterPro; IPR001775; -.
DR   Pfam; PF00263; Bac_GSPproteins; 1.
DR   PRINTS; PR00811; BCTERIALGSPD.
DR   PROSITE; PS00875; T2SP_D; 1.
KW   Transport; Outer membrane; Signal.
FT   SIGNAL        1     25    POTENTIAL.
FT   CHAIN        26    678    GENERAL SECRETION PATHWAY PROTEIN D.
SQ   SEQUENCE  678 AA;  72768 MW;  CB4921C9BAA8438E CRC64;

Query Match             55.2%;    Score 53;  DB 1;  Length 678;
Best Local Similarity   60.0%;    Pred. No. 4;
Matches   12;  Conservative   3;  Mismatches   5;  Indels   0;  Gaps   0;

Qy     2 RVLVLALAVGDGSNLG 21
         :||| |: | :| ||| |||
Db   352 QVLVEAIIVEIADGDGLNLG 371

RESULT   4
VIT2_FUNHE
ID   VIT2_FUNHE       STANDARD;       PRT;       1687 AA.
AC   Q98893;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   VITELLOGENIN II PRECURSOR (VTG II) [CONTAINS: LIPOVITELLIN 1 (LV1);
DE   PHOSVITIN (PV); LIPOVITELLIN 2 (LV2); YP 691].
OS   Fundulus heteroclitus (Killifish) (Mummichog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Fundulidae; Fundulus.
OX   NCBI_TaxID=8078;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 16-35.
RC   TISSUE=Liver;
RA   Lafleur G.J. Jr., Byrne B.M., Haux C., Greenberg R.M., Wallace R.A.;
RT   "Liver-derived cDNAs: vitellogenins and vitelline envelope protein
RT   precursors (choriogenins).";
RL   Int. Symp. Reprod. Physiol. Fish 5:336-338(1995).
CC   -!- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE
CC       SOURCES OF NUTRIENTS DURING EARLY DEVELOPMENT OF OVIPAROUS
CC       ORGANISMS.
CC   -!- TISSUE SPECIFICITY: PRODUCED BY THE LIVER, SECRETED INTO THE BLOOD
CC       AND THEN SEQUESTERED BY RECEPTOR MEDIATED ENDOCYTOSIS INTO GROWING
CC       OOCYTES, WHERE IT IS GENERALLY CLEAVED, GIVING RISE TO THE
CC       RESPECTIVE YOLK COMPONENTS LIPOVITELLINS AND PHOSVITIN.
CC   -!- INDUCTION: SYNTHESIZED IN THE LIVER OF OVIPAROUS VERTEBRATES IN
CC       RESPONSE TO STEROID (ESTROGEN) INDUCTION. STEROID INDUCED
CC       EXPRESSION OF VTG II IS LOWER THAN THAT OF VTG I.
CC   -!- PTM: PHOSVITIN, AN EGG YOLK STORAGE PROTEIN, IS ONE OF THE MOST
CC       HIGHLY PHOSPHORYLATED (10%) PROTEINS IN NATURE (BY SIMILARITY).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U70826; AAB17152.1; -.
DR   InterPro; IPR001747; -.
DR   InterPro; IPR001846; -.
DR   Pfam; PF01347; Vitellogenin_N; 1.
DR   Pfam; PF00094; vwd; 1.
KW   Glycoprotein; Phosphorylation; Storage protein; Signal.
FT   SIGNAL        1     15       POTENTIAL.
FT   CHAIN        16   1687       VITELLOGENIN II.
FT   CHAIN        16      ?       LIPOVITELLIN 1.
FT   CHAIN         ?      ?       PHOSVITIN.
FT   CHAIN         ?   1687       LIPOVITELLIN 2.
FT   DOMAIN     1059   1062       POLY-GLU.
FT   DOMAIN     1088   1169       SER-RICH.
FT   CARBOHYD    941    941       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    945    945       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    954    954       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1004   1004       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1019   1019       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1083   1083       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1142   1142       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1179   1179       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1257   1257       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1292   1292       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1342   1342       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1361   1361       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1366   1366       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1390   1390       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1577   1577       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1655   1655       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   1687 AA;  186005 MW;  4965BB9DBFB4928F CRC64;

Query Match             55.2%;  Score 53;  DB 1;  Length 1687;
Best Local Similarity   75.0%;  Pred. No. 8.2;
Matches  12;  Conservative  1;  Mismatches  3;  Indels  0;  Gaps  0;

Qy          1 MRVLVLALAVALAVGD 16
              |||||||| ||| |:
Db          1 MRVLVLALTVALVAGN 16
```

```
RESULT 5
AT1K_HUMAN
ID   AT1K_HUMAN     STANDARD;      PRT;   945 AA.
AC   O60423;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE IK (EC 3.6.1.-) (FRAGMENT).
GN   FOS37502_2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lanerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA   Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA   Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA   Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA   Kobayashi A., Olsen A.S., Carrano A.V.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   CONCEPTUAL TRANSLATION.
RA   Axelsen K.B.;
RL   Unpublished observations (FEB-2000).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
CC       ATPASES) SUBFAMILY IV.
CC   -!- CAUTION: REF.1 SEQUENCE DIFFERS AT SEVERAL POSITIONS FROM THAT
CC       SHOWN DUE TO DIFFERENCES IN THE PREDICTION OF SPLICE SITES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AC004755; AAC17601.1; ALT_SEQ.
DR   InterPro; IPR001757; -.
DR   InterPro; IPR002048; -.
DR   Pfam; PF00122; E1-E2_ATPase; 2.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW   Multigene family.
FT   NON_TER       1      1
FT   DOMAIN        1    148       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    149    170       POTENTIAL.
FT   DOMAIN      171    197       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    198    219       POTENTIAL.
FT   DOMAIN      220    767       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    768    788       POTENTIAL.
FT   DOMAIN      789    800       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    801    820       POTENTIAL.
FT   DOMAIN      821    850       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    851    872       POTENTIAL.
FT   DOMAIN      873    884       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    885    907       POTENTIAL.
FT   DOMAIN      908    913       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    914    934       POTENTIAL.
FT   DOMAIN      930    933       POLY-THR.
FT   DOMAIN      935    945       EXTRACELLULAR (POTENTIAL).
FT   MOD_RES     262    262       PHOSPHORYLATION (BY SIMILARITY).
FT   METAL       711    711       MAGNESIUM (BY SIMILARITY).
FT   METAL       715    715       MAGNESIUM (BY SIMILARITY).
FT   NON_TER     945    945
SQ   SEQUENCE    945 AA;  107157 MW;  2A28483C2F58D702 CRC64;
```

```
Query Match      46.9%;   Score 45;   DB 1;   Length 945;
Best Local Similarity   41.7%;    Pred. No. 63;
Matches  10;   Conservative  5;   Mismatches  3;   Indels  6;   Gaps  1;

Qy          2 RVLVAL------AVALAVGDGSN 19
                : |::||  | ||:|||:|
Db          691 KALIVALVKKYHQVVTLAIGDGAN 714

RESULT  6
ASM_HUMAN
ID   ASM_HUMAN      STANDARD;      PRT;    629 AA.
AC   P17405; P17406;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   SPHINGOMYELIN PHOSPHODIESTERASE PRECURSOR (EC 3.1.4.12) (ACID
DE   SPHINGOMYELINASE).
GN   SMPD1 OR ASM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91217097; PubMed=1840600;
RA   Schuchman E.H., Suchi M., Takahashi T., Sandhoff K., Desnick R.J.;
RT   "Human acid sphingomyelinase. Isolation, nucleotide sequence and
RT   expression of the full-length and alternatively spliced cDNAs.";
RL   J. Biol. Chem. 266:8531-8539(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93183402; PubMed=1292508;
RA   Newrzella D., Stoffel W.;
RT   "Molecular cloning of the acid sphingomyelinase of the mouse and the
RT   organization and complete nucleotide sequence of the gene.";
RL   Biol. Chem. Hoppe-Seyler 373:1233-1238(1992).
RN   [3]
RP   SEQUENCE OF 128-629 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Fibroblast;
RX   MEDLINE=90060003; PubMed=2555181;
RA   Quintern L.E., Schuchman E.H., Levran O., Suchi M., Ferlinz K.,
RA   Reinke H., Sandhoff K., Desnick R.J.;
RT   "Isolation of cDNA clones encoding human acid sphingomyelinase:
RT   occurrence of alternatively processed transcripts.";
RL   EMBO J. 8:2469-2473(1989).
RN   [4]
RP   VARIANT NPD SER-577.
RX   MEDLINE=92028849; PubMed=1718266;
RA   Ferlinz K., Hurwitz R., Sandhoff K.;
RT   "Molecular basis of acid sphingomyelinase deficiency in a patient
RT   with Niemann-Pick disease type A.";
RL   Biochem. Biophys. Res. Commun. 179:1187-1191(1991).
RN   [5]
RP   VARIANT NPD LEU-496.
RX   MEDLINE=92119449; PubMed=2023926;
RA   Levran O., Desnick R.J., Schuchman E.H.;
RT   "Niemann-Pick disease: a frequent missense mutation in the acid
RT   sphingomyelinase gene of Ashkenazi Jewish type A and B patients.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:3748-3752(1991).
RN   [6]
RP   VARIANT NPD ARG-608 DEL.
RX   MEDLINE=91358737; PubMed=1885770;
RA   Levran O., Desnick R.J., Schuchman E.H.;
RT   "Niemann-Pick type B disease. Identification of a single codon
RT   deletion in the acid sphingomyelinase gene and genotype/phenotype
RT   correlations in type A and B patients.";
RL   J. Clin. Invest. 88:806-810(1991).
RN   [7]
RP   VARIANT NPD PRO-302.
RX   MEDLINE=93004773; PubMed=1391960;
RA   Levran O., Desnick R.J., Schuchman E.H.;
RT   "Identification and expression of a common missense mutation (L302P)
RT   in the acid sphingomyelinase gene of Ashkenazi Jewish type A
RT   Niemann-Pick disease patients.";
RL   Blood 80:2081-2087(1992).
RN   [8]
RP   VARIANT NPD ARG-436.
RX   MEDLINE=92244834; PubMed=1301192;
RA   Takahashi T., Desnick R.J., Takada G., Schuchman E.H.;
RT   "Identification of a missense mutation (S436R) in the acid
RT   sphingomyelinase gene from a Japanese patient with type B
RT   Niemann-Pick disease.";
RL   Hum. Mutat. 1:70-71(1992).
RN   [9]
RP   VARIANTS NPD ARG-242; ILE-382 AND SER-383.
RX   MEDLINE=92316934; PubMed=1618760;
RA   Takahashi T., Suchi M., Desnick R.J., Takada G., Schuchman E.H.;
RT   "Identification and expression of five mutations in the human acid
RT   sphingomyelinase gene causing types A and B Niemann-Pick disease.
RT   Molecular evidence for genetic heterogeneity in the neuronopathic and
RT   non-neuronopathic forms.";
RL   J. Biol. Chem. 267:12552-12558(1992).
RN   [10]
RP   VARIANT NPD GLY-391.
RX   MEDLINE=94328611; PubMed=8051942;
RA   Sperl W., Bart G., Vanier M.T., Christomanou H., Baldissera I.,
RA   Steichensdorf E., Paschke E.;
RT   "A family with visceral course of Niemann-Pick disease, macular halo
RT   syndrome and low sphingomyelin degradation rate.";
RL   J. Inherit. Metab. Dis. 17:93-103(1994).
RN   [11]
RP   VARIANT NPD THR-389.
RX   MEDLINE=96287387; PubMed=8680412;
RA   Schuchman E.H.;
RT   "Two new mutations in the acid sphingomyelinase gene causing type a
RT   Niemann-pick disease: N389T and R441X.";
RL   Hum. Mutat. 6:352-354(1995).
RN   [12]
RP   VARIANT NPD GLN-246.
RX   MEDLINE=96263741; PubMed=8664904;
RA   Ida H., Rennert O.M., Maekawa K., Eto Y.;
RT   "Identification of three novel mutations in the acid
RT   sphingomyelinase gene of Japanese patients with Niemann-Pick disease
RT   type A and B.";
RL   Hum. Mutat. 7:65-67(1996).
CC   -!- FUNCTION: CONVERTS SPHINGOMYELIN TO CERAMIDE. ASM ALSO HAS
CC       PHOSPHOLIPASE C ACTIVITIES TOWARD 1,2-DIACYLGLYCEROLPHOSPHOCHOLINE
CC       AND 1,2-DIACYLGLYCEROLPHOSPHOGLYCEROL.
CC   -!- CATALYTIC ACTIVITY: SPHINGOMYELIN + H(2)O = N-ACYLSPHINGOSINE +
CC       CHOLINE PHOSPHATE.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: LYSOSOMAL.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; ASM-1 (SHOWN HERE)
CC       (FREQUENCY 90%), ASM-2 (FREQUENCY 10%) AND ASM-3 (FREQUENCY <1%);
CC       ARE PRODUCED BY ALTERNATIVE SPLICING. ONLY ASM-1 ENCODES A
CC       CATALYTICALLY ACTIVE ENZYME.
CC   -!- DISEASE: DEFICIENCY IN ASM IS THE CAUSE OF THE NIEMANN-PICK
CC       DISEASE (NPD). SPHINGOMYELIN AND OTHER METABOLICALLY RELATED
CC       LIPIDS ACCUMULATE IN THE LYSOSOMES. PATIENTS MAY SHOW XANTHOMAS,
CC       PIGMENTATION, HEPATOSPLENOMEGALY, LYMPHADENOPATHY AND MENTAL
CC       RETARDATION. THERE ARE TWO MAIN TYPES: TYPE A, IS A RAPIDLY
CC       PROGRESSIVE AND FATAL NEURODEGENERATIVE DISEASE OF INFANCY.
CC       TYPE B, HAS LITTLE IF ANY NEUROLOGIC INVOLVEMENT AND MAY
CC       SURVIVE INTO ADULTHOOD. NPD OCCURS MORE FREQUENTLY AMONG
CC       INDIVIDUALS OF ASHKENAZI JEWISH ANCESTRY THAN IN THE GENERAL
CC       POPULATION.
CC   -!- MISCELLANEOUS: THERE ARE TWO TYPES OF SPHINGOMYELINASES: ASM
CC       (ACID), AND NSM (NEUTRAL).
```

CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M59916; AAA58377.1; -.
DR   EMBL; X63600; CAA45145.1; ALT_SEQ.
DR   EMBL; X52678; CAA36901.1; -.
DR   EMBL; X52679; CAA36902.1; -.
DR   PIR; A39825; A39825.
DR   PIR; S06957; S06957.
DR   PIR; S06958; S06958.
DR   PIR; S27009; S27009.
DR   MIM; 257200; -.
KW   Hydrolase; Glycosidase; Lysosome; Glycoprotein; Alternative splicing;
KW   Signal; Disease mutation.
FT   SIGNAL   1   46
FT   CHAIN   47   629   SPHINGOMYELIN PHOSPHODIESTERASE.
FT   CARBOHYD   86   86   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   175   175   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   335   335   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   395   395   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   503   503   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   520   520   N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   363   374   IGGFYALSPYPG -> YLSSVETQEGKR (IN ISOFORM
FT   ASM 2).
FT   VARSPLIC   375   418   MISSING (IN ISOFORM ASM-2).
FT   VARSPLIC   363   418   MISSING (IN ISOFORM ASM-3).
FT   VARIANT   242   242   G -> R (IN NPD TYPE B).
FT   /FTId=VAR_005058.
FT   VARIANT   246   246   E -> Q (IN NPD TYPE A; 30% RESIDUAL
FT   ACTIVITY).
FT   /FTId=VAR_005059.
FT   VARIANT   302   302   L -> P (IN NPD TYPE A; 23% OF ASHKENAZI
FT   JEWISH TYPE A NPD).
FT   /FTId=VAR_005060.
FT   VARIANT   382   382   M -> I (IN NPD TYPE A).
FT   /FTId=VAR_005061.
FT   VARIANT   383   383   N -> S (IN NPD TYPE B).
FT   /FTId=VAR_005062.
FT   VARIANT   389   389   N -> T (IN NPD TYPE A).
FT   /FTId=VAR_005063.
FT   VARIANT   391   391   W -> G (IN NPD TYPE A/B).
FT   /FTId=VAR_005064.
FT   VARIANT   436   436   S -> R (IN NPD TYPE B).
FT   /FTId=VAR_005065.
FT   VARIANT   496   496   R -> L (IN NPD TYPE A; 32% OF ASHKENAZI
FT   JEWISH TYPE A NPD).
FT   /FTId=VAR_005066.
FT   VARIANT   577   577   G -> S (IN NPD TYPE A).
FT   /FTId=VAR_005067.
FT   VARIANT   608   608   MISSING (IN NPD TYPE B).
FT   /FTId=VAR_005068.
SQ   SEQUENCE   629 AA; 69863 MW; 9C64F04BD2A22A74 CRC64;

Query Match   45.8%; Score 44; DB 1; Length 629;
Best Local Similarity   71.4%; Pred. No. 62;
Matches   10; Conservative   2; Mismatches   2; Indels   2; Gaps   0;

Qy   3 VLVLALAVALAVGD 16
         || |||:|||:| |
Db   36 VLAIALALALALSD 49

RESULT 7
CASB_CAPHI
ID   CASB_CAPHI   STANDARD;   PRT;   222 AA.
AC   P33048;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   BETA CASEIN PRECURSOR.
GN   CSN2.
OS   Capra hircus (Goat).

OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Capra.
OX   NCBI_TaxID=9925;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SAANEN; TISSUE=Blood;
RX   MEDLINE=93077039; PubMed=1446822;
RA   Roberts B., Ditullio P., Vitale J., Hehir K., Gordon K.;
RT   "Cloning of the goat beta-casein-encoding gene and expression in
RT   transgenic mice.";
RL   Gene 121:255-262(1992).
-!- FUNCTION: IMPORTANT ROLE IN DETERMINATION OF THE SURFACE
CC   PROPERTIES OF THE CASEIN MICELLES.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC   -!- SIMILARITY: BELONGS TO THE BETA-CASEIN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M90561; AAA30906.1; -.
DR   EMBL; M90556; AAA30906.1; JOINED.
DR   EMBL; M90557; AAA30906.1; JOINED.
DR   EMBL; M90558; AAA30906.1; JOINED.
DR   EMBL; M90560; AAA30906.1; JOINED.
DR   PIR; JC1384; JC1384.
DR   InterPro; IPR001588; -.
DR   Pfam; PF00363; caseins; 1.
DR   PROSITE; PS00306; CASEIN_ALPHA_BETA; FALSE_NEG.
KW   Milk; Phosphorylation; Glycoprotein; Signal.
FT   SIGNAL   1   15
FT   CHAIN   16   222   BETA CASEIN.
FT   MOD_RES   30   30   PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES   32   32   PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES   33   33   PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES   34   34   PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES   50   50   PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   222 AA; 24865 MW; 96AE1774GA01CD05 CRC64;

Query Match   44.8%; Score 43; DB 1; Length 222;
Best Local Similarity   64.3%; Pred. No. 37;
Matches   9; Conservative   3; Mismatches   3; Indels   2; Gaps   0;

Qy   1 MRVLVLALAVALAV 14
         |:||:|| ||||:
Db   1 MKVLLACLVALAI 14

RESULT 8
APA1_BRARE
ID   APA1_BRARE   STANDARD;   PRT;   262 AA.
AC   O42363;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   APOLIPOPROTEIN A-I PRECURSOR (APO-AI).
GN   APOA.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Rasborinae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=97385153; PubMed=9238027;

RA   Babin P.J., Thisse C., Durliat M., Andre M., Akimenko M.-A.,
RA   Thisse B.;
RT   "Both apolipoprotein E and A-I genes are present in a nonmammalian
RT   vertebrate and are highly expressed during embryonic development.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:8622-8627(1997).
CC   -!- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF
CC   CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING
CC   CHOLESTEROL EFFLUX FROM TISSUES AND BY ACTING AS A COFACTOR FOR
CC   THE LECITHIN CHOLESTEROL ACYLTRANSFERASE (LCAT) (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED IN THE YOLK SYNCYTIAL LAYER
CC   DURING EMBRYONIC (STARTING AT THE GASTRULA STAGE) AND EARLY LARVAL
CC   DEVELOPMENT. AN EXTRAEMBRYONIC STRUCTURE IMPLICATED IN EMBRYONIC
CC   AND LARVAL NUTRITION.
CC   -!- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Y13653; CAA74004.1; -.
DR   ZFIN; ZDB-GENE-990415-14; apoa.
DR   InterPro; IPR000074; -.
DR   Pfam; PF01442; Apolipoprotein; 1.
KW   Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.
FT   SIGNAL   1   18   POTENTIAL.
FT   PROPEP   19   23   POTENTIAL.
FT   CHAIN   24   262   APOLIPOPROTEIN A-I.
FT   DOMAIN   32   63   3 X APPROXIMATE TANDEM REPEATS.
FT   DOMAIN   64   262   10 X APPROXIMATE TANDEM REPEATS.
FT   REPEAT   64   85   1.
FT   REPEAT   87   107   2.
FT   REPEAT   108   118   3 (HALF-LENGTH).
FT   REPEAT   119   140   4.
FT   REPEAT   141   162   5.
FT   REPEAT   163   184   6.
FT   REPEAT   185   206   7.
FT   REPEAT   207   228   8.
FT   REPEAT   229   239   9 (HALF-LENGTH).
FT   REPEAT   240   262   10.
SQ   SEQUENCE   262 AA;   30256 MW; BB839A0A815365B9 CRC64;

Query Match   44.8%; Score 43; DB 1; Length 262;
Best Local Similarity   45.0%; Pred. No. 42;
Matches   9; Conservative   5; Mismatches   5; Indels   0; Gaps   0;

Qy   1 MRVLVLALAVALAVGDGSNL 20
              |:.:||:|||.:||: :||
Db   1 MKFVALALTLLLALGSQANL 20

RESULT   9
BGH3_HUMAN
ID   BGH3_HUMAN   STANDARD;   PRT;   683 AA.
AC   Q15582; O43216; O43217; O43218; O43219; O14471; O14472; O14476;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   TRANSFORMING GROWTH FACTOR-BETA INDUCED PROTEIN IG-H3 PRECURSOR (BETA
DE   IG-H3) (KERATO-EPITHELIN) (RGD-CONTAINING COLLAGEN ASSOCIATED PROTEIN)
DE   (RGD-CAP).
GN   TGFBI OR BIGH3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.

RX   MEDLINE=93000472; PubMed=1388724;
RA   Skonier J., Neubauer M., Madisen L., Bennett K., Plowman G.D.,
RA   Purchio A.F.;
RT   "cDNA cloning and sequence analysis of beta ig-h3, a novel gene
RT   induced in a human adenocarcinoma cell line after treatment with
RT   transforming growth factor-beta.";
RL   DNA Cell Biol. 11:511-522(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kimmerly W., Bondoc M., Cheng J., Connolly K.S., Gunning K.M.,
RA   Davis C.A., Kadner K., Miguel T., Pitluck S., Pollard M., Rojeski H.,
RA   Subramanian S., Martin C.H.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   VARIANTS CORNEAL DYSTROPHIES CYS-124; HIS-124; GLN-555 AND TRP-555.
RX   MEDLINE=97207642; PubMed=9054935;
RA   Munier F.L., Korvatska E., Djemai A., le Paslier D., Zografos L.,
RA   Pescia G., Schorderet D.F.;
RT   "Kerato-epithelin mutations in four 5q31-linked corneal dystrophies.";
RL   Nat. Genet. 15:247-251(1997).
RN   [4]
RP   VARIANTS CORNEAL DYSTROPHIES CYS-124; HIS-124; GLN-555 AND TRP-555.
RX   MEDLINE=98130534; PubMed=9463327;
RA   Korvatska E., Munier F.L., Djemai A., Wang M.X., Frueh B.,
RA   Chiou A.G.-Y., Uffer S., Ballestrazzi E., Braunstein R.E.,
RA   Forster R.K., Culbertson W.W., Boman H., Zografos L., Schorderet D.F.;
RT   "Mutation hot spots in 5q31-linked corneal dystrophies.";
RL   Am. J. Hum. Genet. 62:320-324(1998).
RN   [5]
RP   VARIANT CDL3A THR-501.
RX   MEDLINE=98163459; PubMed=9497262;
RA   Yamamoto S., Okada M., Tsujikawa M., Shimomura Y., Nishida K.,
RA   Inoue Y., Watanabe H., Maeda N., Kurahashi H., Kinoshita S.,
RA   Nakamura Y., Tano Y.;
RT   "A kerato-epithelin (beta-ig-h3) mutation in lattice corneal dystrophy
RT   type IIIA.";
RL   Am. J. Hum. Genet. 62:719-722(1998).
RN   [6]
RP   VARIANT CDRB LEU-124.
RX   MEDLINE=98451378; PubMed=9780098;
RA   Okada M., Yamamoto S., Tsujikawa M., Watanabe H., Inoue Y., Maeda N.,
RA   Shimomura Y., Nishida K., Quantock A.J., Kinoshita S., Tano Y.;
RT   "Two distinct kerato-epithelin mutations in Reis-Bueklers corneal
RT   dystrophy.";
RL   Am. J. Ophthalmol. 126:535-542(1998).
RN   [7]
RP   VARIANT CDL1 ARG-527.
RX   MEDLINE=99013426; PubMed=9799082;
RA   Fujiki K., Hotta Y., Nakayasu K., Yokoyama T., Takano T.,
RA   Yamaguchi T., Kanai A.;
RT   "A new L527R mutation of the betaIGH3 gene in patients with lattice
RT   corneal dystrophy with deep stromal opacities.";
RL   Hum. Genet. 103:286-289(1998).
RN   [8]
RP   VARIANT CDRB PHE-540 DEL.
RA   Rozzo C., Fossarello M., Galleri G., Sole G., Serru A., Orzalesi N.,
RA   Serra A., Pirastu M.;
RT   "A common beta ig-h3 gene mutation (delta F540) in a large cohort of
RT   Sardinian Reis Bueeklers' corneal dystrophy patients.";
RL   Hum. Mutat. 12:215-216(1998).
CC   -!- FUNCTION: BINDS TO TYPE I, II, AND IV COLLAGENS. THIS ADHESION
CC   PROTEIN MAY PLAY AN IMPORTANT ROLE IN CELL-COLLAGEN INTERACTIONS.
CC   IN CARTILAGE, MAY BE INVOLVED IN ENDOCHONDRAL BONE FORMATION.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR. MAY BE ASSOCIATED BOTH WITH
CC   MICROFIBRILS AND WITH THE CELL SURFACE.
CC   -!- INDUCTION: BY TGF-BETA.
CC   -!- DISEASE: DEFECTS IN TGFBI ARE THE CAUSE OF AT LEAST FIVE AUTOSOMAL
CC   DOMINANT TYPES OF CORNEAL DYSTROPHIES: GRANULAR DYSTROPHY GROENOUW
CC   TYPE I (CDGG1), REIS-BUECKLERS CORNEAL DYSTROPHY (CDRB), LATTICE
CC   CORNEAL DYSTROPHY TYPES I AND IIIA (CDL1 AND CDL3A) AND AVELLINO
CC   CORNEAL DYSTROPHY (ACD). CLINICALLY THEY SHOW PROGRESSIVE
CC   OPACIFICATION OF THE CORNEA LEADING TO SEVERE VISUAL HANDICAP.
CC   --

RX   MEDLINE=97267655; PubMed=9112985;
RA   Rawe I.M., Zhan Q., Burrows R., Bennett K., Cintron C.;
RT   "Beta-ig. Molecular cloning and in situ hybridization in corneal
RT   tissues.";
RL   Invest. Ophthalmol. Vis. Sci. 38:893-900(1997).
CC   -!- FUNCTION: BINDS TO TYPE I, II, AND IV.COLLAGENS. THIS ADHESION
CC   PROTEIN MAY PLAY AN IMPORTANT ROLE IN CELL-COLLAGEN INTERACTIONS.
CC   IN CARTILAGE, MAY BE INVOLVED IN ENDOCHONDRAL BONE FORMATION. MAY
CC   PLAY AN IMPORTANT ROLE IN THE MORPHOGENESIS OF CORNEAL TISSUES.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR. MAY BE ASSOCIATED BOTH WITH
CC   MICROFIBRILS AND WITH THE CELL SURFACE (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: LOCATED PRIMARILY IN THE EPITHELIUM OF NORMAL
CC   ADULT CORNEA, IN FETAL STROMAL CELLS, AND BOTH ENDOTHELIUM- AND
CC   STROMA-DERIVED CELLS IN HEALING CORNEAL WOUNDS. NOT EXPRESSED IN
CC   NORMAL ADULT ENDOTHELIUM AND STROMA.
CC   -!- INDUCTION: BY TGF-BETA.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U66205; AAB07015.1; ALT_FRAME.
KW   Extracellular matrix; Signal; Repeat; Cell adhesion.
FT   SIGNAL   1   23   BY SIMILARITY.
FT   CHAIN   24   683   TRANSFORMING GROWTH FACTOR-BETA INDUCED
FT   PROTEIN IG-H3.
FT   DOMAIN   103   236   BIGH3 1.
FT   DOMAIN   240   371   BIGH3 2.
FT   DOMAIN   375   498   BIGH3 3.
FT   DOMAIN   502   632   BIGH3 4.
FT   SITE   641   643   CELL ATTACHMENT SITE (POTENTIAL).
SQ   SEQUENCE   683 AA; 74684 MW; 4548520497548CD6 CRC64;

Query Match   44.8%; Score 43; DB 1; Length 683;
Best Local Similarity   55.6%; Pred. No. 90;
  Matches   10; Conservative   4; Mismatches   4; Indels   0; Gaps   0;

Qy   3 VLVLALAVAVGDGSNL 20
              | :||||:|||:| :|
Db   5 VRLLALALALGPAATL 22

RESULT 11
AT1B_MOUSE
ID   AT1B_MOUSE   STANDARD;   PRT;   1148 AA.
AC   P98200;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE IB (EC 3.6.1.-).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR;
RX   MEDLINE=20473714; PubMed=11015572;
RA   Halleck M.S., Lawler J.F. Jr., Blackshaw S., Gao L., Nagarajan P.,
RA   Hacker C., Pyle S., Newman J.T., Nakanishi Y., Ando H., Weinstock D.,
RA   Williamson P., Schlegel R.A.;
RT   "Differential expression of putative transbilayer amphipath
RT   transporters.";
RL   Physiol. Genomics 1:139-150(1999).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: FOUND IN TESTIS, HEART AND BRAIN. MOST
CC   ABUNDANT IN TESTIS. ALSO DETECTED IN FETAL TISSUES.

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M77349; AAA61163.1; -.
DR   EMBL; AC004503; AAC08849.1; -.
DR   EMBL; AC005219; AAC24944.1; -.
DR   EMBL; AF035626; AAB86695.1; -.
DR   EMBL; AF035627; AAB88698.1; -.
DR   EMBL; AF035628; AAB88696.1; -.
DR   EMBL; AF035629; AAB88697.1; -.
DR   MIM; 601692; -.
DR   MIM; 121900; -.
DR   MIM; 122200; -.
KW   Extracellular matrix; Signal; Repeat; Cell adhesion; Disease mutation.
FT   SIGNAL   1   15   POTENTIAL.
FT   CHAIN   16   683   TRANSFORMING GROWTH FACTOR-BETA INDUCED
FT   PROTEIN IG-H3.
FT   DOMAIN   103   236   BIGH3 1.
FT   DOMAIN   240   371   BIGH3 2.
FT   DOMAIN   375   498   BIGH3 3.
FT   DOMAIN   502   632   BIGH3 4.
FT   SITE   642   644   CELL ATTACHMENT SITE (POTENTIAL).
FT   VARIANT   124   124   R -> C (IN CDL1).
FT   /FTId=VAR_005076.
FT   VARIANT   124   124   R -> H (IN ACD).
FT   /FTId=VAR_005077.
FT   VARIANT   124   124   R -> L (IN CDRB).
FT   /FTId=VAR_005078.
FT   VARIANT   501   501   P -> T (IN CDL3A).
FT   /FTId=VAR_005079.
FT   VARIANT   527   527   L -> R (IN CDL1, LATE-ONSET).
FT   /FTId=VAR_005080.
FT   VARIANT   540   540   MISSING (IN CDRB).
FT   /FTId=VAR_005081.
FT   VARIANT   555   555   R -> Q (IN CDRB).
FT   /FTId=VAR_005082.
FT   VARIANT   555   555   R -> W (IN CDGG1).
FT   /FTId=VAR_005083.
SQ   SEQUENCE   683 AA; 74680 MW; 40FDC8A71EBB3D00 CRC64;

Query Match   44.8%; Score 43; DB 1; Length 683;
Best Local Similarity   55.6%; Pred. No. 90;
  Matches   10; Conservative   4; Mismatches   4; Indels   0; Gaps   0;

Qy   3 VLVLALAVAVGDGSNL 20
              | :||||:|||:| :|
Db   5 VRLLALALALGPAATL 22

RESULT 10
BGH3_RABIT
ID   BGH3_RABIT   STANDARD;   PRT;   683 AA.
AC   Q95215;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   TRANSFORMING GROWTH FACTOR-BETA INDUCED PROTEIN IG-H3 PRECURSOR (BETA
DE   IG-H3) (KERATO-EPITHELIN) (RGD-CONTAINING: COLLAGEN ASSOCIATED PROTEIN)
DE   (RGD-CAP).
GN   TGFBI.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW ZEALAND WHITE; TISSUE=Cornea;

CC   -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
CC   ATPASES). SUBFAMILY IV.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AF156550; AAF09448.1; -.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW   Multigene family.
FT   DOMAIN   1   44   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   45   66   POTENTIAL.
FT   DOMAIN   67   71   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   72   94   POTENTIAL.
FT   DOMAIN   95   276   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   277   298   POTENTIAL.
FT   DOMAIN   299   323   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   324   345   POTENTIAL.
FT   DOMAIN   346   837   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   838   858   POTENTIAL.
FT   DOMAIN   859   870   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   871   890   POTENTIAL.
FT   DOMAIN   891   920   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   921   942   POTENTIAL.
FT   DOMAIN   943   956   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   957   979   POTENTIAL.
FT   DOMAIN   980   985   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   986   1006   POTENTIAL.
FT   DOMAIN   1007   1024   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1025   1049   POTENTIAL.
FT   DOMAIN   1050   1148   CYTOPLASMIC (POTENTIAL).
FT   MOD_RES   388   388   PHOSPHORYLATION (BY SIMILARITY).
FT   METAL   781   781   MAGNESIUM (BY SIMILARITY).
FT   METAL   785   785   MAGNESIUM (BY SIMILARITY).
SQ   SEQUENCE   1148 AA; 129416 MW; A668D6F343CAAAB8 CRC64;

Query Match   44.8%; Score 43; DB 1; Length 1148;
Best Local Similarity   63.6%; Pred. No. 1.4e+02;
  Matches   7; Conservative   3; Mismatches   3; Indels   1; Gaps   0;

Qy   9 AVALAVGDGSN 19
           l: ll:lll:l
Db   774 AITLAIGDGAN 784

RESULT 12
ATIC_HUMAN
ID   ATIC_HUMAN   STANDARD;   PRT;   1251 AA.
AC   O43520;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE IC (EC 3.6.1.-) (FAMILIAL
DE   INTRAHEPATIC CHOLESTASIS TYPE 1).
GN   FIC1 OR PFIC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANTS PFIC1 AND BRIC.
RC   TISSUE=Liver, and intestine;
RX   MEDLINE=98160178; PubMed=9500542;
RA   Bull L.N., van Eijk M.J.T., Pawlikowska L., DeYoung J.A., Juijn J.A.,
RA   Liao M., Lomri N., Berger R., Scharschmidt B.F.,
RA   Knisely A.S., Houwen R.H.J., Freimer N.B.;
RT   "A gene encoding a P-type ATPase mutated in two forms of hereditary

---

RT   cholestasis.";
RL   Nat. Genet. 18:219-223(1998).
RN   [2]
RP   SEQUENCE OF 388-661 FROM N.A.
RC   TISSUE=Colon tumor;
RX   MEDLINE=98217376; PubMed=9548971;
RA   Halleck M.S., Pradhan D., Blackman C., Berkes C., Williamson P.,
RA   Schlegel R.A.;
RT   "Multiple members of a third subfamily of P-type ATPases identified by
RT   genomic sequences and ESTs.";
RL   Genome Res. 8:354-361(1998).
RN   [3]
RP   VARIANT BRIC THR-661.
RX   MEDLINE=99119414; PubMed=9918928;
RA   Tygstrup N., Steig B.A., Juijn J.A., Bull L.N., Houwen R.H.J.;
RT   "Recurrent familial intrahepatic cholestasis in the Faeroe Islands.
RT   Phenotypic heterogeneity but genetic homogeneity.";
RL   Hepatology 29:506-508(1999).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE TRANSPORT OF AMINOPHOSPHOLIPIDS
CC   FROM THE OUTER TO THE INNER LEAFLET OF VARIOUS MEMBRANES AND THE
CC   MAINTENANCE OF ASYMMETRIC DISTRIBUTION OF PHOSPHOLIPIDS IN THE
CC   CANICULAR MEMBRANE. MAY HAVE A ROLE IN TRANSPORT OF BILE ACIDS
CC   INTO THE CANALICULUS, UPTAKE OF BILE ACIDS FROM INTESTINAL
CC   CONTENTS INTO INTESTINAL MUCOSA OR BOTH.
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: FOUND IN MOST TISSUES EXCEPT BRAIN AND
CC   SKELETAL MUSCLE. MOST ABUNDANT IN PANCREAS AND SMALL INTESTINE.
CC   -!- DISEASE: DEFECTS IN FIC1 ARE THE CAUSE OF PROGRESSIVE FAMILIAL
CC   INTRAHEPATIC CHOLESTASIS TYPE 1 (PFIC1 OR BYLER DISEASE), AN
CC   AUTOSOMAL RECESSIVE DISORDER, CHARACTERIZED BY EARLY INFANCY
CC   CHOLESTASIS, THAT MAY BE INITIALLY EPISODIC BUT PROGRESSES TO
CC   MALNUTRITION, GROWTH RETARDATION AND END-STAGE LIVER DISEASE
CC   BEFORE ADULTHOOD.
CC   -!- DISEASE: DEFECTS IN FIC1 ARE THE CAUSE OF BENIGN RECURRENT
CC   INTRAHEPATIC CHOLESTASIS (BRIC OR SUMMERSKILL SYNDROME), AN
CC   AUTOSOMAL RECESSIVE DISORDER, CHARACTERIZED BY INTERMITTENT
CC   EPISODES OF CHOLESTASIS WITHOUT EXTRAHEPATIC BILE DUCT
CC   OBSTRUCTION, WITH INITIAL ELEVATION OF SERUM BILE ACIDS, FOLLOWED
CC   BY CHOLESTATIC JAUNDICE WHICH GENERALLY SPONTANEOUSLY RESOLVES
CC   AFTER PERIODS OF WEEKS TO MONTHS.
CC   -!- DISEASE: DEFECTS IN FIC1 COULD BE ASSOCIATED WITH RECURRENT
CC   INTRAHEPATIC CHOLESTASIS OF PREGNANCY (ICP OR RICP PREGNANCY-
CC   RELATED CHOLESTASIS), CHARACTERIZED BY GENERALIZED ITCHING, WITH
CC   OR WITHOUT JAUNDICE, ABSENCE OF BILIARY COLIC, ABSENCE OF JAUNDICE
CC   OR PRURITUS BETWEEN PREGNANCIES, AND ABSENCE OF CHRONIC LIVER
CC   DISEASE. APPEARS DURING PREGNANCY (PRIMARILY IN THE THIRD
CC   TRIMESTER), AND DISAPPEARS SHORTLY POSTPARTUM. MAY BE A FEMALE-
CC   LIMITED AUTOSOMAL DOMINANT INHERITANCE.
CC   -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
CC   ATPASES). SUBFAMILY IV.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AF038007; AAC63461.1; -.
DR   EMBL; AF032442; AAC04328.1; -.
DR   MIM; 602397; -.
DR   MIM; 211600; -.
DR   MIM; 243300; -.
DR   MIM; 147480; -.
DR   InterPro; IPR001757; -.
DR   Pfam; PF00122; E1-E2_ATPase; 3.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW   Disease mutation; Multigene family.
FT   DOMAIN   1   108   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   109   130   POTENTIAL.

```
FT   DOMAIN      131    136       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    137    156       POTENTIAL.
FT   DOMAIN      157    340       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    341    362       POTENTIAL.
FT   DOMAIN      363    389       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    390    411       POTENTIAL.
FT   DOMAIN      412    949       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    950    970       POTENTIAL.
FT   DOMAIN      971    982       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    983   1002       POTENTIAL.
FT   DOMAIN     1003   1032       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1033   1054       POTENTIAL.
FT   DOMAIN     1055   1068       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN     1069   1091       POTENTIAL.
FT   TRANSMEM   1092   1097       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     1098   1118       POTENTIAL.
FT   DOMAIN     1119   1138       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1139   1163       POTENTIAL.
FT   DOMAIN     1164   1251       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES     454    454       PHOSPHORYLATION (BY SIMILARITY).
FT   METAL       893    893       MAGNESIUM (BY SIMILARITY).
FT   METAL       897    897       MAGNESIUM (BY SIMILARITY).
FT   VARIANT     288    288       L -> S (IN PFIC1).
FT                               /FTId=VAR_008809.
FT   VARIANT     308    308       G -> V (IN PFIC1).
FT                               /FTId=VAR_008810.
FT   VARIANT     645    699       MISSING (IN PFIC1).
FT                               /FTId=VAR_008811.
FT   VARIANT     661    661       I -> T (IN BRIC).
FT                               /FTId=VAR_008812.
FT   VARIANT     892    892       G -> R (IN PFIC1).
FT                               /FTId=VAR_008813.
FT   VARIANT     795    797       MISSING (IN BRIC).
FT                               /FTId=VAR_008814.
SQ   SEQUENCE   1251 AA;  143725 MW;  271EFE24EDA6E144 CRC64;

Query Match            44.8%;  Score 43;  DB 1;  Length 1251;
Best Local Similarity  63.6%;  Pred. No. 1.5e+02;
Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      9  AVALAVGDGSN  19
           |:||:|||:|
Db    886  AITLAIGDGAN  896

RESULT 13
YJBE_ECOLI
ID   YJBE_ECOLI     STANDARD;      PRT;    80 AA.
AC   P32686; Q923E1;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL 7.4 KDA PROTEIN IN PGI-XYLE INTERGENIC REGION PRECURSOR.
GN   YJBE OR B4026.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94089392; PubMed=8265357;
RA   Blattner F.R., Burland V.D., Plunkett G. III; Sofia H.J.,
RA   Daniels D.L.,
RT   "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT   region from 89.2 to 92.8 minutes.";
RL   Nucleic Acids Res. 21:5408-5417(1993).
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U00006; AAC43120.1; -.
DR   EMBL; AE000476; AAC76996.1; -.
DR   EcoGene; EG11923; yjbE.
KW   Hypothetical protein; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22     80       HYPOTHETICAL PROTEIN YJBE.
SQ   SEQUENCE     80 AA;  7386 MW;  A34F87939E46272 CRC64;

Query Match            44.3%;  Score 42.5;  DB 1;  Length 80;
Best Local Similarity  61.1%;  Pred. No. 19;
Matches   11;  Conservative    1;  Mismatches    3;  Indels    3;  Gaps    1;

Qy      7  ALAVALAV--GDGSNLG  21
           |:||||  |||||
Db     53  AVGVALAATGGGDGSNTG  70

RESULT 14
PBP_MANSE
ID   PBP_MANSE      STANDARD;      PRT;   168 AA.
AC   P18959;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PHEROMONE-BINDING PROTEIN PRECURSOR (PBP).
OS   Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Sphingoidea; Sphingidae; Sphinginae; Manduca.
OX   NCBI_TaxID=7130;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 27-56.
RX   MEDLINE=89071794; PubMed=3200861;
RA   Gyoergyi T.K., Roby-Shemkovitz A.J., Lerner M.R.;
RT   "Characterization and cDNA cloning of the pheromone-binding protein
RT   from the tobacco hornworm, Manduca sexta: a tissue-specific
RT   developmentally regulated protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:9851-9855(1988).
RN   [2]
RP   SEQUENCE OF 27-61.
RX   MEDLINE=91186129; PubMed=2010751;
RA   Vogt R.G., Prestwich G.D., Lerner M.R.;
RT   "Odorant-binding-protein subfamilies associate with distinct classes
RT   of olfactory receptor neurons in insects.";
RL   J. Neurobiol. 22:74-84(1991).
CC   -!- FUNCTION: THIS MAJOR SOLUBLE PROTEIN IN OLFACTORY SENSILLA OF MALE
CC       MOTHS MIGHT SERVE TO SOLUBILIZE THE EXTREMELY HYDROPHOBIC
CC       PHEROMONE MOLECULES AND TO TRANSPORT PHEROMONE THROUGH THE AQUEOUS
CC       LYMPH TO RECEPTORS LOCATED ON OLFACTORY CILIA. PBP IS ALSO FOUND
CC       IN SENSILLA FROM FEMALE M.SEXTA ANTENNAE.
CC   -!- SUBUNIT: HOMODIMER (PROBABLE).
CC   -!- TISSUE SPECIFICITY: ANTENNA.
CC   -!- DEVELOPMENTAL STAGE: ITS SYNTHESIS OCCURS AROUND THE TIME OF
CC       ECLOSION.
CC   -!- SIMILARITY: BELONGS TO THE PBP/GOBP FAMILY.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M21797; AAA29325.1; -.
DR   EMBL; M21798; AAA29326.1; -.
DR   PIR; A31770; A31770.
DR   InterPro; IPR000746; -.
```

Left column:

```
DR   Pfam; PF01395; PBP_GOBP; 1.
DR   PRINTS; PR00484; PBPGOBP.
KW   Pheromone response; Transport; Signal.
FT   SIGNAL         1     26
FT   CHAIN         27    168       PHEROMONE-BINDING PROTEIN.
FT   VARIANT       29     29       D -> E.
FT   VARIANT      120    120       I -> V.
SQ   SEQUENCE    168 AA;  18516 MW;  0FC1F18D1908ADF1 CRC64;

Query Match             43.8%;    Score 42;  DB 1;  Length 168;
Best Local Similarity   56.2%;    Pred. No. 40;
Matches    9;  Conservative    4;  Mismatches    4;  Indels    3;  Gaps    0;

Qy      1 MRVLVLALAVAVGD 16
          |:| |:|: :|||:
Db      8 MKVAVVAIVVYLAVGN 23


RESULT 15
APA1_SALSA
ID   APA1_SALSA     STANDARD;      PRT;   258 AA.
AC   P27007;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   APOLIPOPROTEIN A-I PRECURSOR (APO-AI).
OS   Salmo salar (Atlantic salmon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX   NCBI_TaxID=8030;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92009208; PubMed=1916288;
RA   Powell R., Higgins D.G., Wolff J., Byrnes L., Stack M., Sharp P.M.,
RA   Gannon F.;
RT   "The salmon gene encoding apolipoprotein A-I: cDNA sequence, tissue
RT   expression and evolution.";
RL   Gene 104:155-161(1991).
CC   -!- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF
CC       CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING
CC       CHOLESTEROL EFFLUX FROM TISSUES AND BY ACTING AS A COFACTOR FOR
CC       THE LECITHIN CHOLESTEROL ACYLTRANSFERASE (LCAT).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: MAJOR PROTEIN OF PLASMA HDL, ALSO FOUND IN
CC       CHYLOMICRONS. EXPRESSED IN LIVER, INTESTINE, AND MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X52237; CAA36482.1; -.
DR   PIR; JH0472; JH0472.
DR   PIR; S26810; S26810.
DR   InterPro; IPR000074; -.
DR   Pfam; PF01442; Apolipoprotein; 1.
KW   Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.
FT   SIGNAL         1     18       POTENTIAL.
FT   PROPEP        19     31       POTENTIAL.
FT   CHAIN         32    258       APOLIPOPROTEIN A-I.
FT   DOMAIN        32     63       3 X APPROXIMATE TANDEM REPEATS.
FT   DOMAIN        64    258       10 X APPROXIMATE TANDEM REPEATS.
FT   REPEAT        64     85       1.
FT   REPEAT        86    106       2.
FT   REPEAT       107    117       3 (HALF-LENGTH).
FT   REPEAT       118    139       4.
```

Right column:

```
FT   REPEAT       140    161       5.
FT   REPEAT       162    183       6.
FT   REPEAT       184    205       7.
FT   REPEAT       206    227       8.
FT   REPEAT       228    238       9.
FT   REPEAT       239    258       10.
SQ   SEQUENCE    258 AA;  29448 MW;  44B6C7E82D8179F3 CRC64;

Query Match             43.8%;    Score 42;  DB 1;  Length 258;
Best Local Similarity   60.0%;    Pred. No. 57;
Matches    9;  Conservative    2;  Mismatches    2;

Qy      1 MRVLVLALAVALAVG 15
          |: ||||| :| ||
Db      1 MKFLVLALTILLAAG 15


Search completed: May  3, 2001, 15:52:54
Job time: 73 sec
```

OM protein - protein search, using sw model

```
            GenCore version 4.5
    Copyright (c) 1993 - 2000 Compugen Ltd.
```

Run on: May 3, 2001, 15:51:41 ; Search time 18.59 Seconds
                                   (without alignments)
                    64.574 Million cell updates/sec

Title: US-09-426-776-10
Perfect score: 96
Sequence: 1 MRVLVLALAVGDGSNLG 21

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_0401:*
```
 1:  /SIDS6/gcqdata/geneseq/geneseqp/AA1980.DAT:*
 2:  /SIDS6/gcqdata/geneseq/geneseqp/AA1981.DAT:*
 3:  /SIDS6/gcqdata/geneseq/geneseqp/AA1982.DAT:*
 4:  /SIDS6/gcqdata/geneseq/geneseqp/AA1983.DAT:*
 5:  /SIDS6/gcqdata/geneseq/geneseqp/AA1984.DAT:*
 6:  /SIDS6/gcqdata/geneseq/geneseqp/AA1985.DAT:*
 7:  /SIDS6/gcqdata/geneseq/geneseqp/AA1986.DAT:*
 8:  /SIDS6/gcqdata/geneseq/geneseqp/AA1987.DAT:*
 9:  /SIDS6/gcqdata/geneseq/geneseqp/AA1988.DAT:*
10:  /SIDS6/gcqdata/geneseq/geneseqp/AA1989.DAT:*
11:  /SIDS6/gcqdata/geneseq/geneseqp/AA1990.DAT:*
12:  /SIDS6/gcqdata/geneseq/geneseqp/AA1991.DAT:*
13:  /SIDS6/gcqdata/geneseq/geneseqp/AA1992.DAT:*
14:  /SIDS6/gcqdata/geneseq/geneseqp/AA1993.DAT:*
15:  /SIDS6/gcqdata/geneseq/geneseqp/AA1994.DAT:*
16:  /SIDS6/gcqdata/geneseq/geneseqp/AA1995.DAT:*
17:  /SIDS6/gcqdata/geneseq/geneseqp/AA1996.DAT:*
18:  /SIDS6/gcqdata/geneseq/geneseqp/AA1997.DAT:*
19:  /SIDS6/gcqdata/geneseq/geneseqp/AA1998.DAT:*
20:  /SIDS6/gcqdata/geneseq/geneseqp/AA1999.DAT:*
21:  /SIDS6/gcqdata/geneseq/geneseqp/AA2000.DAT:*
22:  /SIDS6/gcqdata/geneseq/geneseqp/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	88	91.7	21	21	Y92778	O. aureus vitellog
2	88	91.7	34	21	Y92780	Vtgss-CAT fusion p
3	88	91.7	38	21	Y92782	Vtgss-EGFP fusion
4	88	91.7	51	21	Y92779	Vtgss-CrFCES fusio
5	88	91.7	51	21	Y92783	Vtgss-beta-lactama
6	45	46.9	313	21	G08134	Arabidopsis thalia
7	45	46.9	313	21	G14879	Arabidopsis thalia
8	45	46.9	313	21	G49241	Arabidopsis thalia
9	45	46.9	313	21	G49246	Arabidopsis thalia
10	45	46.9	325	21	G08133	Arabidopsis thalia
11	45	46.9	325	21	G14878	Arabidopsis thalia
12	45	46.9	325	21	G49240	Arabidopsis thalia
13	45	46.9	325	21	G49245	Arabidopsis thalia
14	45	46.9	345	21	G49244	Arabidopsis thalia
15	45	46.9	349	21	G08132	Arabidopsis thalia
16	45	46.9	353	21	G14877	Arabidopsis thalia
17	45	46.9	353	21	G49239	Arabidopsis thalia
18	45	45.8	908	21	B42511	Human ORFX ORF2275
19	44	45.8	381	18	W35283	Human acid sphingo
20	44	45.8	628	14	R30644	DeltaR608 ASM. Ho
21	44	45.8	628	18	W35262	Human acid sphingo
22	44	45.8	629	14	R30642	ASM protein. Homo
23	44	45.8	629	14	R30643	R496L ASM. Homo s
24	44	45.8	629	18	R30645	L302P ASM. Homo s
25	44	45.8	629	18	W35260	Human acid sphingo
26	44	45.8	629	18	W35261	Human acid sphingo
27	44	45.8	629	18	W35263	Human acid sphingo
28	44	45.8	1000	21	B41414	Human ORFX ORF178
29	44	45.8	1027	14	R42203	Protein L. Peptoc
30	44	45.8	1027	14	R43699	Protein L. Peptoc
31	43	44.8	222	17	W00679	Beta-casein. Capr
32	43	44.8	423	21	Y45093	Mouse lymphoid der
33	43	44.8	466	21	Y95728	Cosmid CHRIM5 enco
34	43	44.8	683	14	R40386	betaIG-H3 protein.
35	43	44.8	683	16	R80573	Human beta-IG-H3 (
36	43	44.8	683	21	B11897	Human colon tumour
37	43	44.8	731	21	B41613	Human ORFX ORF1377
38	43	44.8	887	19	W77409	Mycobacterium Kans
39	43	44.8	1251	20	Y06516	Human cholestatin.
40	42.5	44.3	186	14	R32428	Wheat germ aggluti
41	41	42.7	157	21	G54665	Arabidopsis thalia
42	41	42.7	188	21	G18321	Arabidopsis thalia
43	41	42.7	225	14	R32019	Beta-casein. Homo
44	41	42.7	225	14	R33451	Sequence encoded b
45	41	42.7	225	19	W53664	Human beta-casein.

ALIGNMENTS

```
RESULT   1
Y92778
ID   Y92778 standard; Peptide; 21 AA.
XX
AC   Y92778;
XX
DT   29-AUG-2000  (first entry)
XX
DE   O. aureus vitellogenin secretory signal sequence.
XX
KW   Vtgss; vitellogenin; secretory signal sequence; gene expression;
KW   oestrogen receptor binding protein; systemic circulation.
XX
OS   Oreochromis aureus.
XX
FH   Key             Location/Qualifiers
FT   Cleavage-site   15..16
XX
PN   WO20002366-A1.
XX
PD   11-MAY-2000.
XX
PF   29-OCT-1999;    99WO-SG00108.
XX
PR   30-OCT-1998;    98US-0106426.
PR   26-OCT-1999;    99US-0426776.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
PA   (LAMT/) LAM T J.
XX
PI   Ding JL, Tan NS,  Ho B;
XX
DR   WPI; 2000-365615/31.
DR   N-PSDB; A28495.
```

XX
PT   Isolated nucleic acid for assaying for heterologous gene expression,
PT   detecting presence of compound that binds to estrogen receptor or
PT   producing desired protein from host cell comprises nucleotide sequence
PT   encoding secretory signal sequence
XX
PS   Claim 1; Page 36; 73pp; English.
XX
CC   This sequence is that of a piscine, Oreochromis aureus, vitellogenin
CC   secretory sequence (Vtgss). This and variants that comprise conservative
CC   replacements that retain the biological activities of directing secretion
CC   of a fusion protein from a cell and cleavage of the secretory signal
CC   sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC   fused to either a reporter protein or a lipopolysaccharide-binding
CC   protein coding sequence. The isolated nucleic acid is useful for assaying
CC   for heterologous gene expression, detecting the presence of a compound
CC   that binds to an oestrogen receptor in a sample or producing a desired
CC   protein from a host cell. It can also be used in a method for obtaining
CC   systemic circulation of a desired protein in a transgenic or chimeric
CC   host organism.
XX
SQ   Sequence   21 AA;

Query Match   91.7%;   Score 88; DB 21; Length 21;
Best Local Similarity 95.2%;   Pred. No. 3.4e-07;
Matches   20; Conservative   0; Mismatches 1; Indels   1; Gaps   0;

Qy   1 MRVLVLALAVALAVGDGSNLG 21
          ||||||||||||||||| |||
Db   1 mrvlvlalavalavgdqsnlg 21

RESULT 2
Y92780
ID Y92780 standard; Protein; 34 AA.
XX
AC Y92780;
XX
DT 29-AUG-2000 (first entry)
XX
DE Vtgss-CAT fusion protein (partial).
XX
KW Vtgss; vitellogenin; secretory signal sequence; gene expression; CAT.
XX
OS Chimeric - Oreochromis aureus.
OS Chimeric - Synthetic.
XX
FH Key   Location/Qualifiers
FT Peptide   1..21
FT   /label= secretory_signal_sequence
FT   /note= "O. aureus"
FT Cleavage-site 15..16
FT Protein   22..34
FT   /note= "CAT N-terminal"
XX
PN WO200026366-A1.
XX
PD 11-MAY-2000.
XX
PF 29-OCT-1999;   99WO-SG00108.
XX
PR 30-OCT-1998;   98US-0106426.
PR 26-OCT-1999;   99US-0426776.
XX
PA (UYSI-) UNIV SINGAPORE NAT.
PA (LAMT/) LAM T J.
XX
PI Ding JL, Tan NS, Ho B;
XX
DR WPI; 2000-365615/31.
DR N-PSDB; A28499.

XX
PT   Isolated nucleic acid for assaying for heterologous gene expression,
PT   detecting presence of compound that binds to estrogen receptor or
PT   producing desired protein from host cell comprises nucleotide sequence
PT   encoding secretory signal sequence
XX
PS   Example 2; Fig 5B; 73pp; English.
XX
CC   A reporter CAT system that uses the piscine, Oreochromis aureus,
CC   vitellogenin secretory sequence (Vtgss), designated psp-VtgCAT was
CC   constructed. A28499 and A28722 comprise the 5'- and 3'-ends of the
CC   construct insert. Vtgss and variants that comprise conservative
CC   replacements that retain the biological activities of directing secretion
CC   of a fusion protein from a cell and cleavage of the secretory signal
CC   sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC   fused to either a reporter protein or a lipopolysaccharide-binding
CC   protein coding sequence. The isolated nucleic acid is useful for assaying
CC   for heterologous gene expression, detecting the presence of a compound
CC   that binds to an oestrogen receptor in a sample or producing a desired
CC   protein from a host cell. It can also be used in a method for obtaining
CC   systemic circulation of a desired protein in a transgenic or chimeric
CC   host organism.
XX
SQ   Sequence   34 AA;

Query Match   91.7%;   Score 88; DB 21; Length 34;
Best Local Similarity 95.2%;   Pred. No. 5.8e-07;
Matches   20; Conservative   0; Mismatches 1; Indels   1; Gaps   0;

Qy   1 MRVLVLALAVALAVGDGSNLG 21
          ||||||||||||||||| |||
Db   1 mrvlvlalavalavgdqsnlg 21

RESULT 3
Y92782
ID Y92782 standard; Protein; 38 AA.
XX
AC Y92782;
XX
DT 29-AUG-2000 (first entry)
XX
DE Vtgss-EGFP fusion protein (partial).
XX
KW Vtgss; vitellogenin; secretory signal sequence; gene expression;
KW oestrogen receptor binding protein; systemic circulation; EGFP.
XX
OS Chimeric - Oreochromis aureus.
OS Chimeric - Synthetic.
XX
FH Key   Location/Qualifiers
FT Peptide   1..21
FT   /label= Vtgss
FT Cleavage-site 15..16
FT Protein   28..38
FT   /label= EGFP
XX
PN WO200026366-A1.
XX
PD 11-MAY-2000.
XX
PF 29-OCT-1999;   99WO-SG00108.
XX
PR 30-OCT-1998;   98US-0106426.
PR 26-OCT-1999;   99US-0426776.
XX
PA (UYSI-) UNIV SINGAPORE NAT.
PA (LAMT/) LAM T J.
XX
PI Ding JL, Tan NS, Ho B;
XX
DR WPI; 2000-365615/31.

DR   N-PSDB; A28502.
XX
PT   Isolated nucleic acid for assaying for heterologous gene expression,
PT   detecting presence of compound that binds to estrogen receptor or
PT   producing desired protein from host cell comprises nucleotide sequence
XX
PS   Example 3; Fig 8B; 73pp; English.
XX
CC   A reporter GFP system that uses the piscine, Oreochromis aureus,
CC   vitellogenin secretory sequence (Vtgss), designated pVtgEGFP was
CC   constructed. Vtgss and variants that comprise conservative
CC   replacements that retain the biological activities of directing secretion
CC   of a fusion protein from a cell and cleavage of the secretory signal
CC   sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC   fused to either a reporter protein or a lipopolysaccharide-binding
CC   protein coding sequence. The isolated nucleic acid is useful for assaying
CC   for heterologous gene expression, detecting the presence of a compound
CC   that binds to an oestrogen receptor in a sample or producing a desired
CC   protein from a host cell. It can also be used in a method for obtaining
CC   systemic circulation of a desired protein in a transgenic or chimeric
CC   host organism.
XX
SQ   Sequence   38 AA;

Query Match   91.7%; Score 88; DB 21; Length 38;
Best Local Similarity   95.2%; Pred. No. 6.5e-07;
Matches   20; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 MRVLVLALAVALAVGDGSNLG 21
              |||||||||||||||||| ||||
Db   1 mrvlvlalavalavgdqsnlg 21

RESULT   4
Y92779
ID   Y92779 standard; Protein; 51 AA.
XX
AC   Y92779;
XX
DT   29-AUG-2000 (first entry)
XX
DE   Vtgss-CrFCES fusion protein.
XX
KW   Vtgss; vitellogenin; secretory signal sequence; gene expression;
KW   oestrogen receptor binding protein; systemic circulation; factor C.
XX
OS   Chimeric - Oreochromis aureus.
OS   Chimeric - Carcinoscorpius rotundicauda.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..21
FT   /label= secretory_signal_sequence
FT   /note= "from O. aureus"
FT   Cleavage-site   15..16
FT   Domain   22..51
FT   /label= LPS-binding_domain
FT   /note= "from C. rotundicauda Factor C"
XX
PN   WO200026366-A1.
XX
PD   11-MAY-2000.
XX
PF   29-OCT-1999; 99WO-SG00108.
XX
PR   30-OCT-1998; 98US-0106426.
PR   26-OCT-1999; 99US-0426776.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
PA   (LAMT/) LAM T J.
XX
PI   Ding JL, Tan NS, Ho B;

XX
DR   WPI; 2000-365615/31.
DR   N-PSDB; A28496.
XX
PT   Isolated nucleic acid for assaying for heterologous gene expression,
PT   detecting presence of compound that binds to estrogen receptor or
PT   producing desired protein from host cell comprises nucleotide sequence
PT   encoding secretory signal sequence
XX
PS   Example 1; Fig 2A; 73pp; English.
XX
CC   The piscine, Oreochromis aureus, vitellogenin secretory sequence (Vtgss)
CC   was fused upstream to the EcoRI-SalI cDNA fragment encoding the
CC   lipopolysaccharide-binding domain of Carcinoscorpius rotundicauda Factor
CC   C CrFCES for expression and secretion of recombinant ES protein from
CC   Drosophila cells. Vtgss and variants that comprise conservative
CC   replacements that retain the biological activities of directing secretion
CC   of a fusion protein from a cell and cleavage of the secretory signal
CC   sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC   fused to either a reporter protein or a lipopolysaccharide-binding
CC   protein coding sequence. The isolated nucleic acid is useful for assaying
CC   for heterologous gene expression, detecting the presence of a compound
CC   that binds to an oestrogen receptor in a sample or producing a desired
CC   protein from a host cell. It can also be used in a method for obtaining
CC   systemic circulation of a desired protein in a transgenic or chimeric
CC   host organism.
XX
SQ   Sequence   51 AA;

Query Match   91.7%; Score 88; DB 21; Length 51;
Best Local Similarity   95.2%; Pred. No. 9e-07;
Matches   20; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 MRVLVLALAVALAVGDGSNLG 21
              |||||||||||||||||| ||||
Db   1 mrvlvlalavalavgdqsnlg 21

RESULT   5
Y92783
ID   Y92783 standard; Protein; 51 AA.
XX
AC   Y92783;
XX
DT   29-AUG-2000 (first entry)
XX
DE   Vtgss-beta-lactamase fusion protein (partial).
XX
KW   Vtgss; vitellogenin; secretory signal sequence; gene expression;
KW   oestrogen receptor binding protein; systemic circulation;
KW   beta-lactamase.
XX
OS   Chimeric - Oreochromis aureus.
OS   Chimeric - Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..21
FT   /label= Vtgss
FT   Cleavage-site   15..16
FT   Protein   22..51
FT   /note= "beta-lactamase mature protein"
XX
PN   WO200026366-A1.
XX
PD   11-MAY-2000.
XX
PF   29-OCT-1999; 99WO-SG00108.
XX
PR   30-OCT-1998; 98US-0106426.
PR   26-OCT-1999; 99US-0426776.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.

```
PA   (LAMT/) LAM T J.
XX
PI   Ding JL, Tan NS,  Ho B;
XX
DR   WPI; 2000-365615/31.
DR   N-PSDB; A28507.
XX
PT   Isolated nucleic acid for assaying for heterologous gene expression,
PT   detecting presence of compound that binds to estrogen receptor or
PT   producing desired protein from host cell comprises nucleotide sequence
PT   encoding secretory signal sequence
XX
PS   Example 6; Fig 14A; 73pp; English.
XX
CC   A reporter beta-lactamase system that uses the Oreochromis aureus,
CC   vitellogenin secretory sequence (Vtgss), designated pBADVtgblactKana was
CC   constructed. Vtgss and variants that comprise conservative
CC   replacements that retain the biological activities of directing secretion
CC   of a fusion protein from a cell and cleavage of the secretory signal
CC   sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC   fused to either a reporter protein or a lipopolysaccharide-binding
CC   protein coding sequence. The isolated nucleic acid is useful for assaying
CC   for heterologous gene expression, detecting the presence of a compound
CC   that binds to an oestrogen receptor in a sample or producing a desired
CC   protein from a host cell. It can also be used in a method for obtaining
CC   systemic circulation of a desired protein in a transgenic or chimeric
CC   host organism.
XX
SQ   Sequence  51 AA;

Query Match      91.7%;     Score 88;  DB 21;  Length 51;
Best Local Similarity  95.2%;   Pred. No. 9e-07;
Matches  20;  Conservative  0;  Mismatches  0;  Indels  1;  Indels  0;  Gaps  0;

Qy    1 MRVLVLALAVALAVGDGSNLG 21
         |||||||||||||| ||||
Db    1 mrvlvlalavavgdqsnlg 21


RESULT  6
G08134
ID   G08134 standard; Protein; 313 AA.
XX
AC   G08134;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 5552.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   18-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   08-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   10-JUN-1999;   99US-0139119.
PR   14-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139463.
PR   17-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0142055.
PR   19-JUL-1999;   99US-0142390.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
```

```
PR  20-JUL-1999;  99US-0144352.
PR  20-JUL-1999;  99US-0144632.
PR  20-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0145086.
PR  21-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.
PR  22-JUL-1999;  99US-0145089.
PR  22-JUL-1999;  99US-0145192.
PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  23-JUL-1999;  99US-0145224.
PR  26-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  27-JUL-1999;  99US-0145919.
PR  28-JUL-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  02-AUG-1999;  99US-0146388.
PR  02-AUG-1999;  99US-0146389.
PR  03-AUG-1999;  99US-0147038.
PR  04-AUG-1999;  99US-0147204.
PR  04-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.
PR  05-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  06-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  09-AUG-1999;  99US-0147935.
PR  10-AUG-1999;  99US-0148171.
PR  11-AUG-1999;  99US-0148319.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  13-AUG-1999;  99US-0148684.
PR  16-AUG-1999;  99US-0149368.
PR  17-AUG-1999;  99US-0149175.
PR  18-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  20-AUG-1999;  99US-0149929.
PR  23-AUG-1999;  99US-0149902.
PR  23-AUG-1999;  99US-0149930.
PR  25-AUG-1999;  99US-0150566.
PR  26-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  27-AUG-1999;  99US-0151080.
PR  30-AUG-1999;  99US-0151303.
PR  31-AUG-1999;  99US-0151438.
PR  01-SEP-1999;  99US-0151930.
PR  07-SEP-1999;  99US-0152363.
PR  10-SEP-1999;  99US-0153070.
PR  13-SEP-1999;  99US-0153758.
PR  15-SEP-1999;  99US-0154018.
PR  16-SEP-1999;  99US-0154039.
PR  20-SEP-1999;  99US-0154779.
PR  22-SEP-1999;  99US-0155139.
PR  23-SEP-1999;  99US-0155486.
PR  24-SEP-1999;  99US-0155659.
PR  28-SEP-1999;  99US-0156458.
PR  29-SEP-1999;  99US-0156596.
PR  04-OCT-1999;  99US-0157117.
PR  05-OCT-1999;  99US-0157753.
PR  06-OCT-1999;  99US-0157865.
PR  07-OCT-1999;  99US-0158029.
PR  08-OCT-1999;  99US-0158232.
PR  12-OCT-1999;  99US-0158369.
PR  13-OCT-1999;  99US-0159293.
PR  13-OCT-1999;  99US-0159294.
PR  13-OCT-1999;  99US-0159295.
PR  14-OCT-1999;  99US-0159329.
PR  14-OCT-1999;  99US-0159330.

PR  14-OCT-1999;  99US-0159331.
PR  14-OCT-1999;  99US-0159637.
PR  18-OCT-1999;  99US-0159638.
PR  21-OCT-1999;  99US-0155584.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160767.
PR  21-OCT-1999;  99US-0160768.
PR  21-OCT-1999;  99US-0160770.
PR  21-OCT-1999;  99US-0160814.
PR  21-OCT-1999;  99US-0160815.
PR  22-OCT-1999;  99US-0160980.
PR  22-OCT-1999;  99US-0160981.
PR  22-OCT-1999;  99US-0160989.
PR  25-OCT-1999;  99US-0161404.
PR  25-OCT-1999;  99US-0161405.
PR  25-OCT-1999;  99US-0161406.
PR  26-OCT-1999;  99US-0161359.
PR  26-OCT-1999;  99US-0161360.
PR  26-OCT-1999;  99US-0161361.
PR  28-OCT-1999;  99US-0161920.
PR  28-OCT-1999;  99US-0161992.
PR  28-OCT-1999;  99US-0161993.
PR  29-OCT-1999;  99US-0162142.


Query Match           46.9%;  Score 45; DB 21;  Length 313;
Best Local Similarity 56.2%;  Pred. No. 25;
Matches   9;  Conservative  4;  Mismatches  4;  Indels  0;  Gaps  0;

Qy    4 LVLALAVALAVGDGSN 19
        |::  || |:||:|||
Db  133 llimaavllsvgegsn 148


RESULT  7
G14879
ID   G14879 standard; Protein; 313 AA.
XX
AC   G14879;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 14907.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   23-APR-1999;  99US-0130891.
PR   28-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
```

```
PR  05-MAY-1999;  99US-0132485.
PR  06-MAY-1999;  99US-0132486.
PR  06-MAY-1999;  99US-0132487.
PR  07-MAY-1999;  99US-0132863.
PR  11-MAY-1999;  99US-0134256.
PR  14-MAY-1999;  99US-0134218.
PR  14-MAY-1999;  99US-0134219.
PR  14-MAY-1999;  99US-0134221.
PR  14-MAY-1999;  99US-0134370.
PR  18-MAY-1999;  99US-0134768.
PR  19-MAY-1999;  99US-0134941.
PR  20-MAY-1999;  99US-0135124.
PR  21-MAY-1999;  99US-0135353.
PR  24-MAY-1999;  99US-0135629.
PR  25-MAY-1999;  99US-0136021.
PR  27-MAY-1999;  99US-0136392.
PR  28-MAY-1999;  99US-0136782.
PR  01-JUN-1999;  99US-0137222.
PR  03-JUN-1999;  99US-0137528.
PR  04-JUN-1999;  99US-0137502.
PR  07-JUN-1999;  99US-0137724.
PR  08-JUN-1999;  99US-0138094.
PR  10-JUN-1999;  99US-0138540.
PR  10-JUN-1999;  99US-0138847.
PR  14-JUN-1999;  99US-0139119.
PR  16-JUN-1999;  99US-0139452.
PR  16-JUN-1999;  99US-0139453.
PR  17-JUN-1999;  99US-0139492.
PR  18-JUN-1999;  99US-0139454.
PR  18-JUN-1999;  99US-0139455.
PR  18-JUN-1999;  99US-0139456.
PR  18-JUN-1999;  99US-0139457.
PR  18-JUN-1999;  99US-0139458.
PR  18-JUN-1999;  99US-0139459.
PR  18-JUN-1999;  99US-0139460.
PR  18-JUN-1999;  99US-0139461.
PR  18-JUN-1999;  99US-0139462.
PR  18-JUN-1999;  99US-0139463.
PR  18-JUN-1999;  99US-0139750.
PR  21-JUN-1999;  99US-0139763.
PR  22-JUN-1999;  99US-0139817.
PR  22-JUN-1999;  99US-0139899.
PR  23-JUN-1999;  99US-0140353.
PR  23-JUN-1999;  99US-0140354.
PR  24-JUN-1999;  99US-0140695.
PR  28-JUN-1999;  99US-0140823.
PR  29-JUN-1999;  99US-0140991.
PR  30-JUN-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  01-JUL-1999;  99US-0142154.
PR  02-JUL-1999;  99US-0142055.
PR  06-JUL-1999;  99US-0142390.
PR  08-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144331.
PR  16-JUL-1999;  99US-0144332.
PR  16-JUL-1999;  99US-0144005.
PR  19-JUL-1999;  99US-0144085.
PR  19-JUL-1999;  99US-0144334.
PR  19-JUL-1999;  99US-0144335.
PR  19-JUL-1999;  99US-0144086.
PR  20-JUL-1999;  99US-0144352.
PR  20-JUL-1999;  99US-0144632.
PR  20-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0145086.
PR  21-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145085.

PR  22-JUL-1999;  99US-0145087.
PR  22-JUL-1999;  99US-0145089.
PR  22-JUL-1999;  99US-0145192.
PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  23-JUL-1999;  99US-0145224.
PR  26-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  27-JUL-1999;  99US-0145919.
PR  28-JUL-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  02-AUG-1999;  99US-0146388.
PR  02-AUG-1999;  99US-0146389.
PR  03-AUG-1999;  99US-0147038.
PR  04-AUG-1999;  99US-0147204.
PR  04-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.
PR  05-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  06-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  09-AUG-1999;  99US-0147935.
PR  10-AUG-1999;  99US-0148171.
PR  11-AUG-1999;  99US-0148319.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  13-AUG-1999;  99US-0148684.
PR  16-AUG-1999;  99US-0149368.
PR  17-AUG-1999;  99US-0149175.
PR  18-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  20-AUG-1999;  99US-0149929.
PR  23-AUG-1999;  99US-0149902.
PR  23-AUG-1999;  99US-0149930.
PR  25-AUG-1999;  99US-0150566.
PR  26-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  27-AUG-1999;  99US-0151080.
PR  30-AUG-1999;  99US-0151303.
PR  31-AUG-1999;  99US-0151438.
PR  01-SEP-1999;  99US-0151930.
PR  07-SEP-1999;  99US-0152363.
PR  10-SEP-1999;  99US-0153070.
PR  13-SEP-1999;  99US-0153758.
PR  15-SEP-1999;  99US-0154018.
PR  16-SEP-1999;  99US-0154039.
PR  20-SEP-1999;  99US-0154779.
PR  22-SEP-1999;  99US-0155139.
PR  23-SEP-1999;  99US-0155486.
PR  24-SEP-1999;  99US-0155659.
PR  28-SEP-1999;  99US-0156458.
PR  29-SEP-1999;  99US-0156596.
PR  04-OCT-1999;  99US-0157117.
PR  05-OCT-1999;  99US-0157753.
PR  06-OCT-1999;  99US-0157865.
PR  07-OCT-1999;  99US-0158029.
PR  08-OCT-1999;  99US-0158232.
PR  12-OCT-1999;  99US-0158369.
PR  13-OCT-1999;  99US-0159293.
PR  13-OCT-1999;  99US-0159294.
PR  13-OCT-1999;  99US-0159295.
PR  14-OCT-1999;  99US-0159329.
PR  14-OCT-1999;  99US-0159330.
PR  14-OCT-1999;  99US-0159331.
PR  14-OCT-1999;  99US-0159637.
PR  18-OCT-1999;  99US-0159638.
PR  21-OCT-1999;  99US-0159584.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160767.
PR  21-OCT-1999;  99US-0160768.
```

```
PR   21-OCT-1999;  99US-0160770.                PR   14-MAY-1999;  99US-0134221.
PR   21-OCT-1999;  99US-0160814.                PR   14-MAY-1999;  99US-0134370.
PR   21-OCT-1999;  99US-0160815.                PR   18-MAY-1999;  99US-0134768.
PR   22-OCT-1999;  99US-0160980.                PR   19-MAY-1999;  99US-0134941.
PR   22-OCT-1999;  99US-0160981.                PR   20-MAY-1999;  99US-0135124.
PR   22-OCT-1999;  99US-0160989.                PR   21-MAY-1999;  99US-0135353.
PR   25-OCT-1999;  99US-0161404.                PR   24-MAY-1999;  99US-0135629.
PR   25-OCT-1999;  99US-0161405.                PR   25-MAY-1999;  99US-0136021.
PR   25-OCT-1999;  99US-0161406.                PR   27-MAY-1999;  99US-0136392.
PR   26-OCT-1999;  99US-0161359.                PR   28-MAY-1999;  99US-0136782.
PR   26-OCT-1999;  99US-0161360.                PR   01-JUN-1999;  99US-0137222.
PR   26-OCT-1999;  99US-0161361.                PR   03-JUN-1999;  99US-0137528.
PR   28-OCT-1999;  99US-0161920.                PR   04-JUN-1999;  99US-0137502.
PR   28-OCT-1999;  99US-0161992.                PR   07-JUN-1999;  99US-0137724.
PR   28-OCT-1999;  99US-0161993.                PR   08-JUN-1999;  99US-0138094.
PR   29-OCT-1999;  99US-0162142.                PR   10-JUN-1999;  99US-0138540.
                                                PR   10-JUN-1999;  99US-0138847.
Query Match     46.9%;   Score 45;  DB 21;  Length 313;
                                                PR   14-JUN-1999;  99US-0139119.
Best Local Similarity  56.2%;  Pred. NO. 25;    PR   16-JUN-1999;  99US-0139452.
Matches   9;  Conservative   4;  Mismatches   4;  Indels   3;  Gaps   0;
                                                PR   16-JUN-1999;  99US-0139453.
                                                PR   17-JUN-1999;  99US-0139492.
Qy    4 LVLALAVALAVGDGSN 19                      PR   18-JUN-1999;  99US-0139454.
        |:: ||:||:|||                            PR   18-JUN-1999;  99US-0139455.
Db  133 llimaavllsvgegsn 148                     PR   18-JUN-1999;  99US-0139456.
                                                PR   18-JUN-1999;  99US-0139457.
                                                PR   18-JUN-1999;  99US-0139458.
RESULT   8                                       PR   18-JUN-1999;  99US-0139459.
G49241                                           PR   18-JUN-1999;  99US-0139460.
ID   G49241 standard; Protein; 313 AA.           PR   18-JUN-1999;  99US-0139461.
XX                                               PR   18-JUN-1999;  99US-0139462.
AC   G49241;                                     PR   18-JUN-1999;  99US-0139463.
XX                                               PR   18-JUN-1999;  99US-0139750.
DT   18-OCT-2000  (first entry)                  PR   18-JUN-1999;  99US-0139763.
XX                                               PR   21-JUN-1999;  99US-0139817.
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62272.
                                                PR   22-JUN-1999;  99US-0139899.
XX                                               PR   23-JUN-1999;  99US-0140353.
KW   Protein identification; signal transduction pathway; metabolic pathway;
                                                PR   23-JUN-1999;  99US-0140354.
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
                                                PR   24-JUN-1999;  99US-0140695.
KW   termination sequence.                       PR   28-JUN-1999;  99US-0140823.
XX                                               PR   29-JUN-1999;  99US-0140991.
OS   Arabidopsis thaliana.                       PR   30-JUN-1999;  99US-0141287.
XX                                               PR   01-JUL-1999;  99US-0141842.
PN   EP1033405-A2.                               PR   01-JUL-1999;  99US-0142154.
XX                                               PR   02-JUL-1999;  99US-0142055.
PD   06-SEP-2000.                                PR   06-JUL-1999;  99US-0142390.
XX                                               PR   08-JUL-1999;  99US-0142803.
PF   25-FEB-2000;  2000EP-0301439.               PR   09-JUL-1999;  99US-0142920.
XX                                               PR   12-JUL-1999;  99US-0142977.
PR   25-FEB-1999;  99US-0121825.                 PR   13-JUL-1999;  99US-0143542.
PR   05-MAR-1999;  99US-0123180.                 PR   14-JUL-1999;  99US-0143624.
PR   09-MAR-1999;  99US-0123548.                 PR   15-JUL-1999;  99US-0144005.
PR   23-MAR-1999;  99US-0125788.                 PR   16-JUL-1999;  99US-0144085.
PR   25-MAR-1999;  99US-0126264.                 PR   16-JUL-1999;  99US-0144086.
PR   29-MAR-1999;  99US-0126785.                 PR   19-JUL-1999;  99US-0144325.
PR   01-APR-1999;  99US-0127462.                 PR   19-JUL-1999;  99US-0144331.
PR   06-APR-1999;  99US-0128234.                 PR   19-JUL-1999;  99US-0144332.
PR   08-APR-1999;  99US-0128714.                 PR   19-JUL-1999;  99US-0144333.
PR   16-APR-1999;  99US-0129845.                 PR   19-JUL-1999;  99US-0144334.
PR   19-APR-1999;  99US-0130077.                 PR   19-JUL-1999;  99US-0144335.
PR   21-APR-1999;  99US-0130449.                 PR   20-JUL-1999;  99US-0144352.
PR   23-APR-1999;  99US-0130510.                 PR   20-JUL-1999;  99US-0144632.
PR   28-APR-1999;  99US-0130891.                 PR   20-JUL-1999;  99US-0144884.
PR   30-APR-1999;  99US-0131449.                 PR   21-JUL-1999;  99US-0144814.
PR   04-MAY-1999;  99US-0132048.                 PR   21-JUL-1999;  99US-0145086.
PR   05-MAY-1999;  99US-0132407.                 PR   21-JUL-1999;  99US-0145088.
PR   06-MAY-1999;  99US-0132484.                 PR   22-JUL-1999;  99US-0145085.
PR   06-MAY-1999;  99US-0132485.                 PR   22-JUL-1999;  99US-0145087.
PR   07-MAY-1999;  99US-0132486.                 PR   22-JUL-1999;  99US-0145089.
PR   11-MAY-1999;  99US-0132487.                 PR   22-JUL-1999;  99US-0145192.
PR   14-MAY-1999;  99US-0132863.                 PR   23-JUL-1999;  99US-0145145.
PR   14-MAY-1999;  99US-0134256.                 PR   23-JUL-1999;  99US-0145218.
         99US-0134218.                           PR   23-JUL-1999;  99US-0145224.
         99US-0134219.                           PR   26-JUL-1999;  99US-0145276.
```

```
PR  27-JUL-1999;  99US-0145913.          PR  25-OCT-1999;  99US-0161405.
PR  27-JUL-1999;  99US-0145918.          PR  25-OCT-1999;  99US-0161406.
PR  27-JUL-1999;  99US-0145919.          PR  26-OCT-1999;  99US-0161359.
PR  28-JUL-1999;  99US-0145951.          PR  26-OCT-1999;  99US-0161360.
PR  02-AUG-1999;  99US-0146386.          PR  26-OCT-1999;  99US-0161361.
PR  02-AUG-1999;  99US-0146388.          PR  28-OCT-1999;  99US-0161920.
PR  03-AUG-1999;  99US-0146389.          PR  28-OCT-1999;  99US-0161992.
PR  04-AUG-1999;  99US-0147038.          PR  28-OCT-1999;  99US-0161993.
PR  04-AUG-1999;  99US-0147204.          PR  29-OCT-1999;  99US-0162142.
PR  05-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.          Query Match                Score 45;  DB 21;  Length 313;
PR  06-AUG-1999;  99US-0147260.          Best Local Similarity  46.9%;     Pred. No.  25;
PR  06-AUG-1999;  99US-0147303.          Matches   9;  Conservative   4;  Mismatches   4;  Indels   3;  Gaps   0;
PR  09-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.          Qy    4 LVLALAVALAVGDGSN 19
PR  10-AUG-1999;  99US-0147935.                  |::: || |:||:|||
PR  11-AUG-1999;  99US-0148171.          Db  133 llimaavllsvgegsn 148
PR  12-AUG-1999;  99US-0148319.
PR  13-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.          RESULT   9
PR  16-AUG-1999;  99US-0148684.          G49246
PR  17-AUG-1999;  99US-0149368.          ID  G49246 standard; Protein; 313 AA.
PR  18-AUG-1999;  99US-0149175.          XX
PR  20-AUG-1999;  99US-0149426.          AC  G49246;
PR  20-AUG-1999;  99US-0149722.          XX
PR  20-AUG-1999;  99US-0149723.          DT  18-OCT-2000  (first entry)
PR  23-AUG-1999;  99US-0149902.          XX
PR  25-AUG-1999;  99US-0149929.          DE  Arabidopsis thaliana protein fragment SEQ ID NO: 62279.
PR  25-AUG-1999;  99US-0149930.          XX
PR  26-AUG-1999;  99US-0150566.          KW  Protein identification; signal transduction pathway; metabolic pathway;
PR  27-AUG-1999;  99US-0150884.          KW  hybridisation assay; genetic mapping; gene expression control; promoter;
PR  27-AUG-1999;  99US-0151065.          KW  termination sequence.
PR  27-AUG-1999;  99US-0151066.          XX
PR  30-AUG-1999;  99US-0151080.          OS  Arabidopsis thaliana.
PR  31-AUG-1999;  99US-0151303.          XX
PR  01-SEP-1999;  99US-0151438.          PN  EP1033405-A2.
PR  07-SEP-1999;  99US-0151930.          XX
PR  10-SEP-1999;  99US-0152363.          PD  06-SEP-2000.
PR  13-SEP-1999;  99US-0153070.          XX
PR  15-SEP-1999;  99US-0153758.          PF  25-FEB-2000;  2000EP-0301439.
PR  16-SEP-1999;  99US-0154018.          XX
PR  20-SEP-1999;  99US-0154039.          PR  25-FEB-1999;  99US-0121825.
PR  22-SEP-1999;  99US-0154779.          PR  05-MAR-1999;  99US-0123180.
PR  23-SEP-1999;  99US-0155139.          PR  09-MAR-1999;  99US-0123548.
PR  24-SEP-1999;  99US-0155486.          PR  23-MAR-1999;  99US-0125788.
PR  28-SEP-1999;  99US-0155659.          PR  25-MAR-1999;  99US-0126264.
PR  29-SEP-1999;  99US-0156458.          PR  29-MAR-1999;  99US-0126785.
PR  04-OCT-1999;  99US-0156596.          PR  01-APR-1999;  99US-0127462.
PR  06-OCT-1999;  99US-0157117.          PR  06-APR-1999;  99US-0128234.
PR  07-OCT-1999;  99US-0157753.          PR  08-APR-1999;  99US-0128714.
PR  08-OCT-1999;  99US-0157865.          PR  16-APR-1999;  99US-0129845.
PR  12-OCT-1999;  99US-0158029.          PR  19-APR-1999;  99US-0130077.
PR  13-OCT-1999;  99US-0158232.          PR  21-APR-1999;  99US-0130449.
PR  13-OCT-1999;  99US-0158369.          PR  23-APR-1999;  99US-0130510.
PR  14-OCT-1999;  99US-0159293.          PR  28-APR-1999;  99US-0130891.
PR  14-OCT-1999;  99US-0159294.          PR  29-APR-1999;  99US-0131449.
PR  14-OCT-1999;  99US-0159295.          PR  30-APR-1999;  99US-0132048.
PR  14-OCT-1999;  99US-0159329.          PR  04-MAY-1999;  99US-0132407.
PR  18-OCT-1999;  99US-0159330.          PR  05-MAY-1999;  99US-0132484.
PR  21-OCT-1999;  99US-0159331.          PR  06-MAY-1999;  99US-0132485.
PR  21-OCT-1999;  99US-0159637.          PR  05-MAY-1999;  99US-0132486.
PR  21-OCT-1999;  99US-0159638.          PR  07-MAY-1999;  99US-0132487.
PR  21-OCT-1999;  99US-0159584.          PR  11-MAY-1999;  99US-0132863.
PR  22-OCT-1999;  99US-0160741.          PR  14-MAY-1999;  99US-0134256.
PR  22-OCT-1999;  99US-0160767.          PR  14-MAY-1999;  99US-0134218.
PR  25-OCT-1999;  99US-0160768.          PR  14-MAY-1999;  99US-0134219.
PR  ...                99US-0160770.          PR  18-MAY-1999;  99US-0134221.
PR  ...                99US-0160814.          PR  19-MAY-1999;  99US-0134370.
PR  ...                99US-0160815.          PR  20-MAY-1999;  99US-0134768.
PR  ...                99US-0160980.          PR  21-MAY-1999;  99US-0134941.
PR  ...                99US-0160981.          PR  24-MAY-1999;  99US-0135124.
PR  ...                99US-0160989.          PR                     99US-0135353.
PR  ...                99US-0161404.          PR                     99US-0135629.
```

us-09-426-776-10.rag

PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	30-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	31-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	01-SEP-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	07-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	10-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	13-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	15-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	16-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	20-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	22-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	23-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	24-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	28-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	29-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	04-OCT-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	05-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	06-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	07-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	08-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	12-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	13-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	21-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0159584.
PR	22-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	22-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	22-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	25-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	26-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	28-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.

```
PR   28-OCT-1999;    99US-0161993.
PR   29-OCT-1999;    99US-0162142.

Query Match           46.9%;    Score 45;  DB 21;  Length 313;
Best Local Similarity  56.2%;   Pred. No. 25;
Matches    9;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 LVLALAVALAVGDGSN 19
          |:: ||  |:||:|||
Db    133 llimaavllsvgegsn 148


RESULT   10
G08133
ID   G08133 standard; Protein; 325 AA.
XX
AC   G08133;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 5551.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;    99US-0121825.        PR   08-JUN-1999;    99US-0138094.
PR   05-MAR-1999;    99US-0123180.        PR   10-JUN-1999;    99US-0138540.
PR   09-MAR-1999;    99US-0123548.        PR   10-JUN-1999;    99US-0138847.
PR   23-MAR-1999;    99US-0125788.        PR   14-JUN-1999;    99US-0139119.
PR   25-MAR-1999;    99US-0126264.        PR   16-JUN-1999;    99US-0139452.
PR   29-MAR-1999;    99US-0126785.        PR   16-JUN-1999;    99US-0139453.
PR   01-APR-1999;    99US-0127462.        PR   17-JUN-1999;    99US-0139492.
PR   06-APR-1999;    99US-0128234.        PR   18-JUN-1999;    99US-0139454.
PR   08-APR-1999;    99US-0128714.        PR   18-JUN-1999;    99US-0139455.
PR   16-APR-1999;    99US-0129845.        PR   18-JUN-1999;    99US-0139456.
PR   19-APR-1999;    99US-0130077.        PR   18-JUN-1999;    99US-0139457.
PR   21-APR-1999;    99US-0130449.        PR   18-JUN-1999;    99US-0139458.
PR   23-APR-1999;    99US-0130510.        PR   18-JUN-1999;    99US-0139459.
PR   23-APR-1999;    99US-0130891.        PR   18-JUN-1999;    99US-0139460.
PR   28-APR-1999;    99US-0131449.        PR   18-JUN-1999;    99US-0139461.
PR   30-APR-1999;    99US-0132048.        PR   18-JUN-1999;    99US-0139462.
PR   30-APR-1999;    99US-0132407.        PR   18-JUN-1999;    99US-0139463.
PR   04-MAY-1999;    99US-0132484.        PR   18-JUN-1999;    99US-0139750.
PR   05-MAY-1999;    99US-0132485.        PR   18-JUN-1999;    99US-0139763.
PR   06-MAY-1999;    99US-0132486.        PR   21-JUN-1999;    99US-0139817.
PR   06-MAY-1999;    99US-0132487.        PR   22-JUN-1999;    99US-0139899.
PR   07-MAY-1999;    99US-0132863.        PR   23-JUN-1999;    99US-0140353.
PR   11-MAY-1999;    99US-0134256.        PR   23-JUN-1999;    99US-0140354.
PR   14-MAY-1999;    99US-0134218.        PR   24-JUN-1999;    99US-0140695.
PR   14-MAY-1999;    99US-0134219.        PR   28-JUN-1999;    99US-0140823.
PR   14-MAY-1999;    99US-0134221.        PR   29-JUN-1999;    99US-0140991.
PR   14-MAY-1999;    99US-0134370.        PR   30-JUN-1999;    99US-0141287.
PR   18-MAY-1999;    99US-0134768.        PR   01-JUL-1999;    99US-0141842.
PR   19-MAY-1999;    99US-0134941.        PR   01-JUL-1999;    99US-0142154.
PR   20-MAY-1999;    99US-0135124.        PR   02-JUL-1999;    99US-0142055.
PR   21-MAY-1999;    99US-0135353.        PR   06-JUL-1999;    99US-0142390.
PR   24-MAY-1999;    99US-0135629.        PR   08-JUL-1999;    99US-0142803.
PR   25-MAY-1999;    99US-0136021.        PR   09-JUL-1999;    99US-0142920.
PR   27-MAY-1999;    99US-0136392.        PR   12-JUL-1999;    99US-0142977.
PR   28-MAY-1999;    99US-0136782.        PR   13-JUL-1999;    99US-0143542.
PR   01-JUN-1999;    99US-0137222.        PR   14-JUL-1999;    99US-0143624.
PR   03-JUN-1999;    99US-0137528.        PR   15-JUL-1999;    99US-0144005.
PR   04-JUN-1999;    99US-0137502.        PR   16-JUL-1999;    99US-0144085.
PR   07-JUN-1999;    99US-0137724.        PR   16-JUL-1999;    99US-0144086.
                                          PR   19-JUL-1999;    99US-0144325.
                                          PR   19-JUL-1999;    99US-0144331.
                                          PR   19-JUL-1999;    99US-0144332.
                                          PR   19-JUL-1999;    99US-0144333.
                                          PR   19-JUL-1999;    99US-0144334.
                                          PR   19-JUL-1999;    99US-0144335.
                                          PR   20-JUL-1999;    99US-0144352.
                                          PR   20-JUL-1999;    99US-0144632.
                                          PR   21-JUL-1999;    99US-0144884.
                                          PR   21-JUL-1999;    99US-0144814.
                                          PR   21-JUL-1999;    99US-0145086.
                                          PR   22-JUL-1999;    99US-0145088.
                                          PR   22-JUL-1999;    99US-0145085.
                                          PR   22-JUL-1999;    99US-0145087.
                                          PR   22-JUL-1999;    99US-0145089.
                                          PR   23-JUL-1999;    99US-0145192.
                                          PR   23-JUL-1999;    99US-0145145.
                                          PR   23-JUL-1999;    99US-0145218.
                                          PR   26-JUL-1999;    99US-0145224.
                                          PR   27-JUL-1999;    99US-0145276.
                                          PR   27-JUL-1999;    99US-0145913.
                                          PR   28-JUL-1999;    99US-0145918.
                                          PR   28-JUL-1999;    99US-0145919.
                                          PR   02-AUG-1999;    99US-0145951.
                                          PR   02-AUG-1999;    99US-0146386.
                                          PR   02-AUG-1999;    99US-0146388.
                                          PR   03-AUG-1999;    99US-0146389.
                                          PR   04-AUG-1999;    99US-0147038.
                                          PR   04-AUG-1999;    99US-0147204.
                                          PR   05-AUG-1999;    99US-0147302.
                                          PR   05-AUG-1999;    99US-0147192.
                                          PR   06-AUG-1999;    99US-0147260.
                                          PR   06-AUG-1999;    99US-0147303.
                                          PR   06-AUG-1999;    99US-0147416.
```

```
Qy      4 LVLALAVALAVGDGSN 19
          |:: || |:||:|||
Db    145 llimaavllsvgegsn 160
```

RESULT 11
G14878
ID   G14878 standard; Protein; 325 AA.
XX
AC   G14878;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 14906.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   25-MAY-1999; 99US-0136392.
PR   27-MAY-1999; 99US-0136782.
PR   28-MAY-1999; 99US-0137222.
PR   01-JUN-1999; 99US-0137528.
PR   03-JUN-1999; 99US-0137502.
PR   04-JUN-1999; 99US-0137724.
PR   07-JUN-1999; 99US-0138094.
PR   08-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   10-JUN-1999; 99US-0139119.
PR   14-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.
PR   16-JUN-1999; 99US-0139492.
PR   17-JUN-1999; 99US-0139492.
PR   09-AUG-1999; 99US-0147493.
PR   09-AUG-1999; 99US-0147935.
PR   10-AUG-1999; 99US-0148171.
PR   11-AUG-1999; 99US-0148319.
PR   12-AUG-1999; 99US-0148341.
PR   13-AUG-1999; 99US-0148565.
PR   13-AUG-1999; 99US-0148684.
PR   16-AUG-1999; 99US-0149368.
PR   17-AUG-1999; 99US-0149175.
PR   18-AUG-1999; 99US-0149426.
PR   20-AUG-1999; 99US-0149722.
PR   20-AUG-1999; 99US-0149929.
PR   23-AUG-1999; 99US-0149902.
PR   23-AUG-1999; 99US-0149723.
PR   23-AUG-1999; 99US-0149930.
PR   25-AUG-1999; 99US-0150566.
PR   26-AUG-1999; 99US-0150884.
PR   27-AUG-1999; 99US-0151065.
PR   27-AUG-1999; 99US-0151066.
PR   27-AUG-1999; 99US-0151080.
PR   30-AUG-1999; 99US-0151303.
PR   31-AUG-1999; 99US-0151438.
PR   01-SEP-1999; 99US-0151930.
PR   07-SEP-1999; 99US-0152363.
PR   10-SEP-1999; 99US-0153070.
PR   13-SEP-1999; 99US-0153758.
PR   15-SEP-1999; 99US-0154018.
PR   16-SEP-1999; 99US-0154039.
PR   20-SEP-1999; 99US-0154779.
PR   22-SEP-1999; 99US-0155139.
PR   23-SEP-1999; 99US-0155486.
PR   24-SEP-1999; 99US-0155659.
PR   28-SEP-1999; 99US-0156458.
PR   29-SEP-1999; 99US-0156596.
PR   04-OCT-1999; 99US-0157117.
PR   05-OCT-1999; 99US-0157753.
PR   06-OCT-1999; 99US-0157865.
PR   07-OCT-1999; 99US-0158029.
PR   08-OCT-1999; 99US-0158232.
PR   12-OCT-1999; 99US-0158369.
PR   13-OCT-1999; 99US-0159293.
PR   13-OCT-1999; 99US-0159294.
PR   13-OCT-1999; 99US-0159295.
PR   14-OCT-1999; 99US-0159329.
PR   14-OCT-1999; 99US-0159330.
PR   14-OCT-1999; 99US-0159331.
PR   14-OCT-1999; 99US-0159637.
PR   14-OCT-1999; 99US-0159638.
PR   18-OCT-1999; 99US-0159584.
PR   21-OCT-1999; 99US-0160741.
PR   21-OCT-1999; 99US-0160767.
PR   21-OCT-1999; 99US-0160768.
PR   21-OCT-1999; 99US-0160770.
PR   21-OCT-1999; 99US-0160814.
PR   21-OCT-1999; 99US-0160815.
PR   22-OCT-1999; 99US-0160980.
PR   22-OCT-1999; 99US-0160981.
PR   22-OCT-1999; 99US-0160989.
PR   25-OCT-1999; 99US-0161404.
PR   25-OCT-1999; 99US-0161405.
PR   25-OCT-1999; 99US-0161406.
PR   26-OCT-1999; 99US-0161359.
PR   26-OCT-1999; 99US-0161360.
PR   26-OCT-1999; 99US-0161361.
PR   28-OCT-1999; 99US-0161920.
PR   28-OCT-1999; 99US-0161992.
PR   28-OCT-1999; 99US-0161993.
PR   29-OCT-1999; 99US-0162142.
```

Query Match      46.9%;   Score 45;   DB 21;   Length 325;
Best Local Similarity   56.2%;   Pred. No. 26;
Matches    9;   Conservative   4;   Mismatches    4;   Indels    3;   Gaps    0;

| | | | |
|---|---|---|---|
| PR | 18-JUN-1999; | 99US-0139454. | |
| PR | 18-JUN-1999; | 99US-0139455. | |
| PR | 18-JUN-1999; | 99US-0139456. | |
| PR | 18-JUN-1999; | 99US-0139457. | |
| PR | 18-JUN-1999; | 99US-0139458. | |
| PR | 18-JUN-1999; | 99US-0139459. | |
| PR | 18-JUN-1999; | 99US-0139460. | |
| PR | 18-JUN-1999; | 99US-0139461. | |
| PR | 18-JUN-1999; | 99US-0139462. | |
| PR | 18-JUN-1999; | 99US-0139463. | |
| PR | 18-JUN-1999; | 99US-0139750. | |
| PR | 18-JUN-1999; | 99US-0139763. | |
| PR | 21-JUN-1999; | 99US-0139817. | |
| PR | 22-JUN-1999; | 99US-0139899. | |
| PR | 23-JUN-1999; | 99US-0140353. | |
| PR | 23-JUN-1999; | 99US-0140354. | |
| PR | 24-JUN-1999; | 99US-0140695. | |
| PR | 28-JUN-1999; | 99US-0140823. | |
| PR | 29-JUN-1999; | 99US-0140991. | |
| PR | 30-JUN-1999; | 99US-0141287. | |
| PR | 01-JUL-1999; | 99US-0141842. | |
| PR | 01-JUL-1999; | 99US-0142154. | |
| PR | 02-JUL-1999; | 99US-0142055. | |
| PR | 06-JUL-1999; | 99US-0142390. | |
| PR | 08-JUL-1999; | 99US-0142803. | |
| PR | 09-JUL-1999; | 99US-0142920. | |
| PR | 12-JUL-1999; | 99US-0142977. | |
| PR | 13-JUL-1999; | 99US-0143542. | |
| PR | 14-JUL-1999; | 99US-0143624. | |
| PR | 15-JUL-1999; | 99US-0144005. | |
| PR | 16-JUL-1999; | 99US-0144085. | |
| PR | 19-JUL-1999; | 99US-0144086. | |
| PR | 19-JUL-1999; | 99US-0144325. | |
| PR | 19-JUL-1999; | 99US-0144331. | |
| PR | 19-JUL-1999; | 99US-0144332. | |
| PR | 19-JUL-1999; | 99US-0144333. | |
| PR | 19-JUL-1999; | 99US-0144334. | |
| PR | 19-JUL-1999; | 99US-0144335. | |
| PR | 20-JUL-1999; | 99US-0144352. | |
| PR | 20-JUL-1999; | 99US-0144632. | |
| PR | 20-JUL-1999; | 99US-0144884. | |
| PR | 21-JUL-1999; | 99US-0144814. | |
| PR | 21-JUL-1999; | 99US-0145086. | |
| PR | 22-JUL-1999; | 99US-0145085. | |
| PR | 22-JUL-1999; | 99US-0145087. | |
| PR | 22-JUL-1999; | 99US-0145089. | |
| PR | 22-JUL-1999; | 99US-0145192. | |
| PR | 23-JUL-1999; | 99US-0145145. | |
| PR | 23-JUL-1999; | 99US-0145218. | |
| PR | 23-JUL-1999; | 99US-0145224. | |
| PR | 26-JUL-1999; | 99US-0145276. | |
| PR | 27-JUL-1999; | 99US-0145913. | |
| PR | 27-JUL-1999; | 99US-0145918. | |
| PR | 27-JUL-1999; | 99US-0145919. | |
| PR | 28-JUL-1999; | 99US-0145951. | |
| PR | 02-AUG-1999; | 99US-0146386. | |
| PR | 02-AUG-1999; | 99US-0147416. | |
| PR | 02-AUG-1999; | 99US-0147493. | |
| PR | 09-AUG-1999; | 99US-0147935. | |
| PR | 10-AUG-1999; | 99US-0146388. | |
| PR | 11-AUG-1999; | 99US-0146389. | |
| PR | 12-AUG-1999; | 99US-0147038. | |
| PR | 13-AUG-1999; | 99US-0148319. | |
| PR | 13-AUG-1999; | 99US-0148341. | |
| PR | 13-AUG-1999; | 99US-0148565. | |
| PR | 13-AUG-1999; | 99US-0148684. | |

| | | | |
|---|---|---|---|
| PR | 16-AUG-1999; | 99US-0149368. | |
| PR | 17-AUG-1999; | 99US-0149175. | |
| PR | 18-AUG-1999; | 99US-0149426. | |
| PR | 20-AUG-1999; | 99US-0149722. | |
| PR | 20-AUG-1999; | 99US-0149723. | |
| PR | 20-AUG-1999; | 99US-0149929. | |
| PR | 23-AUG-1999; | 99US-0149902. | |
| PR | 23-AUG-1999; | 99US-0149930. | |
| PR | 25-AUG-1999; | 99US-0150566. | |
| PR | 26-AUG-1999; | 99US-0150084. | |
| PR | 27-AUG-1999; | 99US-0151065. | |
| PR | 27-AUG-1999; | 99US-0151066. | |
| PR | 27-AUG-1999; | 99US-0151080. | |
| PR | 30-AUG-1999; | 99US-0151303. | |
| PR | 31-AUG-1999; | 99US-0151438. | |
| PR | 01-SEP-1999; | 99US-0151930. | |
| PR | 01-SEP-1999; | 99US-0153363. | |
| PR | 07-SEP-1999; | 99US-0153070. | |
| PR | 10-SEP-1999; | 99US-0153758. | |
| PR | 13-SEP-1999; | 99US-0154018. | |
| PR | 15-SEP-1999; | 99US-0154039. | |
| PR | 16-SEP-1999; | 99US-0154779. | |
| PR | 20-SEP-1999; | 99US-0155139. | |
| PR | 22-SEP-1999; | 99US-0155486. | |
| PR | 23-SEP-1999; | 99US-0155659. | |
| PR | 24-SEP-1999; | 99US-0156458. | |
| PR | 28-SEP-1999; | 99US-0156596. | |
| PR | 29-SEP-1999; | 99US-0157117. | |
| PR | 29-SEP-1999; | 99US-0157753. | |
| PR | 04-OCT-1999; | 99US-0157865. | |
| PR | 05-OCT-1999; | 99US-0158029. | |
| PR | 06-OCT-1999; | 99US-0158232. | |
| PR | 07-OCT-1999; | 99US-0158369. | |
| PR | 08-OCT-1999; | 99US-0159293. | |
| PR | 12-OCT-1999; | 99US-0159294. | |
| PR | 13-OCT-1999; | 99US-0159295. | |
| PR | 13-OCT-1999; | 99US-0159329. | |
| PR | 13-OCT-1999; | 99US-0159330. | |
| PR | 14-OCT-1999; | 99US-0159331. | |
| PR | 14-OCT-1999; | 99US-0159637. | |
| PR | 14-OCT-1999; | 99US-0159638. | |
| PR | 18-OCT-1999; | 99US-0159584. | |
| PR | 21-OCT-1999; | 99US-0160741. | |
| PR | 21-OCT-1999; | 99US-0160767. | |
| PR | 21-OCT-1999; | 99US-0160768. | |
| PR | 21-OCT-1999; | 99US-0160770. | |
| PR | 21-OCT-1999; | 99US-0160814. | |
| PR | 22-OCT-1999; | 99US-0160815. | |
| PR | 22-OCT-1999; | 99US-0160980. | |
| PR | 22-OCT-1999; | 99US-0160981. | |
| PR | 25-OCT-1999; | 99US-0160989. | |
| PR | 25-OCT-1999; | 99US-0161404. | |
| PR | 26-OCT-1999; | 99US-0161405. | |
| PR | 26-OCT-1999; | 99US-0161406. | |
| PR | 26-OCT-1999; | 99US-0161359. | |
| PR | 28-OCT-1999; | 99US-0161360. | |
| PR | 28-OCT-1999; | 99US-0161361. | |
| PR | 29-OCT-1999; | 99US-0161920. | |
| PR | | 99US-0161992. | |
| PR | | 99US-0161993. | |
| PR | | 99US-0162142. | |

```
Query Match                      46.9%;  Score 45;  DB 21;  Length 325;
Best Local Similarity            56.2%;  Pred. No. 26;
  Matches    9;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      4  LVLALAVALAVGDGSN 19
           |:: || |:||:|||
Db    145  llimaavllsvgegsn 160


RESULT  12
G49240
```

```
ID   G49240 standard; Protein; 325 AA.
XX
AC   G49240;
XX
DT   18-OCT-2000  (first entry)
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62271.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.             PR   18-JUN-1999;  99US-0139461.
PR   05-MAR-1999;  99US-0123180.             PR   18-JUN-1999;  99US-0139462.
PR   09-MAR-1999;  99US-0123548.             PR   18-JUN-1999;  99US-0139463.
PR   23-MAR-1999;  99US-0125788.             PR   18-JUN-1999;  99US-0139750.
PR   25-MAR-1999;  99US-0126264.             PR   19-JUN-1999;  99US-0139763.
PR   29-MAR-1999;  99US-0126785.             PR   21-JUN-1999;  99US-0139817.
PR   01-APR-1999;  99US-0127462.             PR   22-JUN-1999;  99US-0139899.
PR   06-APR-1999;  99US-0128234.             PR   23-JUN-1999;  99US-0140353.
PR   08-APR-1999;  99US-0128714.             PR   23-JUN-1999;  99US-0140354.
PR   16-APR-1999;  99US-0129845.             PR   24-JUN-1999;  99US-0140695.
PR   19-APR-1999;  99US-0130077.             PR   28-JUN-1999;  99US-0140823.
PR   21-APR-1999;  99US-0130449.             PR   29-JUN-1999;  99US-0140991.
PR   23-APR-1999;  99US-0130510.             PR   30-JUN-1999;  99US-0141287.
PR   28-APR-1999;  99US-0130891.             PR   01-JUL-1999;  99US-0141842.
PR   30-APR-1999;  99US-0131449.             PR   01-JUL-1999;  99US-0142154.
PR   30-APR-1999;  99US-0132048.             PR   02-JUL-1999;  99US-0142055.
PR   04-MAY-1999;  99US-0132407.             PR   06-JUL-1999;  99US-0142390.
PR   05-MAY-1999;  99US-0132484.             PR   08-JUL-1999;  99US-0142803.
PR   06-MAY-1999;  99US-0132485.             PR   09-JUL-1999;  99US-0142920.
PR   06-MAY-1999;  99US-0132486.             PR   12-JUL-1999;  99US-0142977.
PR   07-MAY-1999;  99US-0132487.             PR   13-JUL-1999;  99US-0143542.
PR   11-MAY-1999;  99US-0132863.             PR   14-JUL-1999;  99US-0143624.
PR   14-MAY-1999;  99US-0134256.             PR   15-JUL-1999;  99US-0144005.
PR   14-MAY-1999;  99US-0134218.             PR   16-JUL-1999;  99US-0144085.
PR   14-MAY-1999;  99US-0134219.             PR   16-JUL-1999;  99US-0144086.
PR   18-MAY-1999;  99US-0134221.             PR   19-JUL-1999;  99US-0144325.
PR   19-MAY-1999;  99US-0134370.             PR   19-JUL-1999;  99US-0144331.
PR   20-MAY-1999;  99US-0134768.             PR   19-JUL-1999;  99US-0144332.
PR   21-MAY-1999;  99US-0134941.             PR   19-JUL-1999;  99US-0144333.
PR   24-MAY-1999;  99US-0135124.             PR   19-JUL-1999;  99US-0144334.
PR   25-MAY-1999;  99US-0135353.             PR   19-JUL-1999;  99US-0144335.
PR   27-MAY-1999;  99US-0135629.             PR   20-JUL-1999;  99US-0144352.
PR   01-JUN-1999;  99US-0136021.             PR   20-JUL-1999;  99US-0144632.
PR   03-JUN-1999;  99US-0136392.             PR   20-JUL-1999;  99US-0144884.
PR   04-JUN-1999;  99US-0137222.             PR   21-JUL-1999;  99US-0144814.
PR   07-JUN-1999;  99US-0137528.             PR   21-JUL-1999;  99US-0145086.
PR   08-JUN-1999;  99US-0137502.             PR   21-JUL-1999;  99US-0145088.
PR   10-JUN-1999;  99US-0137724.             PR   22-JUL-1999;  99US-0145085.
PR   14-JUN-1999;  99US-0138094.             PR   22-JUL-1999;  99US-0145087.
PR   16-JUN-1999;  99US-0138540.             PR   22-JUL-1999;  99US-0145089.
PR   17-JUN-1999;  99US-0138847.             PR   22-JUL-1999;  99US-0145192.
PR   18-JUN-1999;  99US-0139119.             PR   23-JUL-1999;  99US-0145145.
PR   18-JUN-1999;  99US-0139452.             PR   23-JUL-1999;  99US-0145218.
PR   18-JUN-1999;  99US-0139453.             PR   23-JUL-1999;  99US-0145224.
PR   18-JUN-1999;  99US-0139492.             PR   26-JUL-1999;  99US-0145276.
PR   18-JUN-1999;  99US-0139454.             PR   27-JUL-1999;  99US-0145913.
PR   18-JUN-1999;  99US-0139455.             PR   27-JUL-1999;  99US-0145918.
PR   18-JUN-1999;  99US-0139456.             PR   27-JUL-1999;  99US-0145919.
PR   18-JUN-1999;  99US-0139457.             PR   28-JUL-1999;  99US-0145951.
PR   18-JUN-1999;  99US-0139458.             PR   02-AUG-1999;  99US-0146386.
PR   18-JUN-1999;  99US-0139459.             PR   02-AUG-1999;  99US-0146388.
PR   18-JUN-1999;  99US-0139460.             PR   02-AUG-1999;  99US-0146389.
                                             PR   03-AUG-1999;  99US-0147038.
                                             PR   04-AUG-1999;  99US-0147204.
                                             PR   05-AUG-1999;  99US-0147302.
                                             PR   05-AUG-1999;  99US-0147192.
                                             PR   06-AUG-1999;  99US-0147260.
                                             PR   06-AUG-1999;  99US-0147303.
                                             PR   09-AUG-1999;  99US-0147416.
                                             PR   09-AUG-1999;  99US-0147493.
                                             PR   10-AUG-1999;  99US-0147935.
                                             PR   11-AUG-1999;  99US-0148171.
                                             PR   12-AUG-1999;  99US-0148319.
                                             PR   13-AUG-1999;  99US-0148341.
                                             PR   13-AUG-1999;  99US-0148565.
                                             PR   16-AUG-1999;  99US-0148684.
                                             PR   17-AUG-1999;  99US-0149368.
                                             PR   18-AUG-1999;  99US-0149175.
                                             PR   20-AUG-1999;  99US-0149426.
                                             PR   20-AUG-1999;  99US-0149722.
                                             PR   20-AUG-1999;  99US-0149723.
                                             PR   23-AUG-1999;  99US-0149929.
                                             PR   23-AUG-1999;  99US-0149902.
```

```
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   18-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139919.
PR   16-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139455.
PR   17-JUN-1999; 99US-0139492.
PR   18-JUN-1999; 99US-0139454.
PR   18-JUN-1999; 99US-0139455.
PR   18-JUN-1999; 99US-0139456.
PR   18-JUN-1999; 99US-0139457.
PR   18-JUN-1999; 99US-0139458.
PR   18-JUN-1999; 99US-0139459.
PR   18-JUN-1999; 99US-0139460.
PR   18-JUN-1999; 99US-0139461.
PR   18-JUN-1999; 99US-0139462.
PR   18-JUN-1999; 99US-0139463.
PR   18-JUN-1999; 99US-0139750.
PR   18-JUN-1999; 99US-0139763.
PR   21-JUN-1999; 99US-0139817.
PR   22-JUN-1999; 99US-0139899.

PR   23-AUG-1999; 99US-0149930.
PR   25-AUG-1999; 99US-0150566.
PR   26-AUG-1999; 99US-0150884.
PR   27-AUG-1999; 99US-0151065.
PR   27-AUG-1999; 99US-0151066.
PR   30-AUG-1999; 99US-0151080.
PR   31-AUG-1999; 99US-0151303.
PR   31-AUG-1999; 99US-0151438.
PR   01-SEP-1999; 99US-0151930.
PR   07-SEP-1999; 99US-0152363.
PR   10-SEP-1999; 99US-0153070.
PR   13-SEP-1999; 99US-0153758.
PR   15-SEP-1999; 99US-0154018.
PR   16-SEP-1999; 99US-0154039.
PR   20-SEP-1999; 99US-0154779.
PR   22-SEP-1999; 99US-0155139.
PR   23-SEP-1999; 99US-0155486.
PR   24-SEP-1999; 99US-0155659.
PR   28-SEP-1999; 99US-0156458.
PR   29-SEP-1999; 99US-0156596.
PR   04-OCT-1999; 99US-0157117.
PR   05-OCT-1999; 99US-0157753.
PR   06-OCT-1999; 99US-0157865.
PR   07-OCT-1999; 99US-0158029.
PR   08-OCT-1999; 99US-0158232.
PR   12-OCT-1999; 99US-0158369.
PR   13-OCT-1999; 99US-0159293.
PR   13-OCT-1999; 99US-0159294.
PR   14-OCT-1999; 99US-0159295.
PR   14-OCT-1999; 99US-0159329.
PR   14-OCT-1999; 99US-0159330.
PR   14-OCT-1999; 99US-0159331.
PR   14-OCT-1999; 99US-0159637.
PR   14-OCT-1999; 99US-0159638.
PR   18-OCT-1999; 99US-0159584.
PR   21-OCT-1999; 99US-0160741.
PR   21-OCT-1999; 99US-0160767.
PR   21-OCT-1999; 99US-0160768.
PR   21-OCT-1999; 99US-0160770.
PR   21-OCT-1999; 99US-0160814.
PR   21-OCT-1999; 99US-0160815.
PR   22-OCT-1999; 99US-0160980.
PR   22-OCT-1999; 99US-0160981.
PR   22-OCT-1999; 99US-0160989.
PR   25-OCT-1999; 99US-0161404.
PR   25-OCT-1999; 99US-0161405.
PR   25-OCT-1999; 99US-0161406.
PR   26-OCT-1999; 99US-0161359.
PR   26-OCT-1999; 99US-0161360.
PR   28-OCT-1999; 99US-0161361.
PR   28-OCT-1999; 99US-0161920.
PR   28-OCT-1999; 99US-0161992.
PR   29-OCT-1999; 99US-0161993.
PR   29-OCT-1999; 99US-0162142.
```

```
Query Match             46.9%;  Score 45;  DB 21;  Length 325;
Best Local Similarity   56.2%;  Pred. No. 26;
Matches    9;  Conservative    4;  Mismatches    4;  Indels    3;  Gaps    0;

Qy         4 LVLALAVALAVGDGSN 19
             |:.:: || |:||:|||
Db       145 llimaavllsvgegsn 160


RESULT 13
G49245
ID   G49245 standard; Protein; 325 AA.
XX
AC   G49245;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62278.
```

```
PR  23-JUN-1999;  99US-0140353.
PR  23-JUN-1999;  99US-0140354.
PR  24-JUN-1999;  99US-0140695.
PR  28-JUN-1999;  99US-0140823.
PR  29-JUN-1999;  99US-0140991.
PR  30-JUN-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  01-JUL-1999;  99US-0142154.
PR  02-JUL-1999;  99US-0142055.
PR  06-JUL-1999;  99US-0142390.
PR  08-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144005.
PR  16-JUL-1999;  99US-0144085.
PR  19-JUL-1999;  99US-0144086.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144331.
PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.
PR  19-JUL-1999;  99US-0144334.
PR  20-JUL-1999;  99US-0144335.
PR  20-JUL-1999;  99US-0144352.
PR  20-JUL-1999;  99US-0144632.
PR  20-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0145086.
PR  22-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.
PR  22-JUL-1999;  99US-0145089.
PR  22-JUL-1999;  99US-0145192.
PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  26-JUL-1999;  99US-0145224.
PR  27-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  28-JUL-1999;  99US-0145919.
PR  02-AUG-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  03-AUG-1999;  99US-0146388.
PR  04-AUG-1999;  99US-0146389.
PR  05-AUG-1999;  99US-0147038.
PR  05-AUG-1999;  99US-0147204.
PR  06-AUG-1999;  99US-0147302.
PR  09-AUG-1999;  99US-0147260.
PR  10-AUG-1999;  99US-0147303.
PR  11-AUG-1999;  99US-0147416.
PR  13-AUG-1999;  99US-0147493.
PR  13-AUG-1999;  99US-0147935.
PR  17-AUG-1999;  99US-0148171.
PR  18-AUG-1999;  99US-0148319.
PR  20-AUG-1999;  99US-0148341.
PR  20-AUG-1999;  99US-0148565.
PR  20-AUG-1999;  99US-0148684.
PR  23-AUG-1999;  99US-0149368.
PR  23-AUG-1999;  99US-0149175.
PR  25-AUG-1999;  99US-0149426.
PR  26-AUG-1999;  99US-0149722.
PR  27-AUG-1999;  99US-0149723.
PR  27-AUG-1999;  99US-0149929.
PR  30-AUG-1999;  99US-0149902.
PR             ;  99US-0149930.
PR  25-AUG-1999;  99US-0150566.
PR  26-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  27-AUG-1999;  99US-0151080.
PR  30-AUG-1999;  99US-0151303.

PR  31-AUG-1999;  99US-0151438.
PR  01-SEP-1999;  99US-0151930.
PR  07-SEP-1999;  99US-0152363.
PR  10-SEP-1999;  99US-0153070.
PR  13-SEP-1999;  99US-0153758.
PR  15-SEP-1999;  99US-0154018.
PR  16-SEP-1999;  99US-0154039.
PR  20-SEP-1999;  99US-0154779.
PR  22-SEP-1999;  99US-0155139.
PR  23-SEP-1999;  99US-0155486.
PR  24-SEP-1999;  99US-0155659.
PR  28-SEP-1999;  99US-0156458.
PR  29-SEP-1999;  99US-0156596.
PR  04-OCT-1999;  99US-0157117.
PR  05-OCT-1999;  99US-0157753.
PR  06-OCT-1999;  99US-0157865.
PR  07-OCT-1999;  99US-0158029.
PR  08-OCT-1999;  99US-0158232.
PR  12-OCT-1999;  99US-0158369.
PR  13-OCT-1999;  99US-0159293.
PR  13-OCT-1999;  99US-0159294.
PR  13-OCT-1999;  99US-0159295.
PR  14-OCT-1999;  99US-0159329.
PR  14-OCT-1999;  99US-0159330.
PR  14-OCT-1999;  99US-0159331.
PR  14-OCT-1999;  99US-0159637.
PR  14-OCT-1999;  99US-0159638.
PR  18-OCT-1999;  99US-0159984.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160767.
PR  21-OCT-1999;  99US-0160768.
PR  21-OCT-1999;  99US-0160770.
PR  21-OCT-1999;  99US-0160814.
PR  22-OCT-1999;  99US-0160815.
PR  22-OCT-1999;  99US-0160980.
PR  25-OCT-1999;  99US-0160981.
PR  25-OCT-1999;  99US-0160989.
PR  25-OCT-1999;  99US-0161404.
PR  26-OCT-1999;  99US-0161405.
PR  26-OCT-1999;  99US-0161406.
PR  26-OCT-1999;  99US-0161359.
PR  28-OCT-1999;  99US-0161360.
PR  28-OCT-1999;  99US-0161361.
PR  28-OCT-1999;  99US-0161920.
PR  28-OCT-1999;  99US-0161992.
PR  29-OCT-1999;  99US-0161993.
PR  29-OCT-1999;  99US-0162142.
```

```
Query Match             46.9%;          Score 45;  DB 21;  Length 325;
Best Local Similarity   56.2%;          Pred. No. 26;
Matches    9;  Conservative  9;  Mismatches  4;  Indels  3;  Gaps  0;

Qy      4 LVLALAVALAVGDGSN 19
          |::  ||  |:||:|||
Db    145 llimaavllsvgegsn 160

RESULT 14
G49244
ID   G49244 standard; Protein; 345 AA.
XX
AC   G49244;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62277.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
```

Mon May 7 10:35:59 2001

us-09-426-776-10.rag

PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000;	2000EP-0301439.
XX		
PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	14-JUN-1999;	99US-0138847.
PR	16-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	17-JUN-1999;	99US-0139453.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.

PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	22-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	23-JUL-1999;	99US-0145089.
PR	23-JUL-1999;	99US-0145145.
PR	26-JUL-1999;	99US-0145218.
PR	27-JUL-1999;	99US-0145192.
PR	27-JUL-1999;	99US-0145224.
PR	27-JUL-1999;	99US-0145276.
PR	28-JUL-1999;	99US-0145913.
PR	02-AUG-1999;	99US-0145918.
PR	02-AUG-1999;	99US-0145919.
PR	03-AUG-1999;	99US-0145951.
PR	04-AUG-1999;	99US-0146386.
PR	04-AUG-1999;	99US-0146388.
PR	05-AUG-1999;	99US-0146389.
PR	06-AUG-1999;	99US-0147038.
PR	09-AUG-1999;	99US-0147204.
PR	09-AUG-1999;	99US-0147302.
PR	10-AUG-1999;	99US-0147192.
PR	11-AUG-1999;	99US-0147260.
PR	12-AUG-1999;	99US-0147303.
PR	13-AUG-1999;	99US-0147416.
PR	13-AUG-1999;	99US-0147493.
PR	16-AUG-1999;	99US-0147935.
PR	17-AUG-1999;	99US-0148171.
PR	18-AUG-1999;	99US-0148319.
PR	20-AUG-1999;	99US-0148341.
PR	20-AUG-1999;	99US-0148565.
PR	23-AUG-1999;	99US-0148684.
PR	25-AUG-1999;	99US-0149368.
PR	26-AUG-1999;	99US-0149175.
PR	27-AUG-1999;	99US-0149426.
PR	27-AUG-1999;	99US-0149722.
PR	30-AUG-1999;	99US-0149723.
PR	31-AUG-1999;	99US-0149929.
PR	01-SEP-1999;	99US-0149902.
PR	07-SEP-1999;	99US-0149930.
PR	10-SEP-1999;	99US-0150566.
PR	13-SEP-1999;	99US-0150884.
PR	15-SEP-1999;	99US-0151065.
PR	16-SEP-1999;	99US-0151066.
PR		99US-0151080.
PR		99US-0151303.
PR		99US-0151438.
PR		99US-0151930.
PR		99US-0152363.
PR		99US-0153070.
PR		99US-0153758.
PR		99US-0154018.
PR		99US-0154039.

PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Query Match   46.9%;   Score 45; DB 21; Length 345;
Best Local Similarity   56.2%;   Pred. No. 27;
Matches   9; Conservative   4; Mismatches   3; Indels   0; Gaps   0;

Qy   4 LVLALAVALAVGDCGSN 19
         |::: || |:|:|||
Db 165 llimaavllsvgegsn 180

RESULT 15
G08132
ID   G08132 standard; Protein; 349 AA.
XX
AC   G08132;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 5550.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.

PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   21-JUN-1999;   99US-0139763.
PR   22-JUN-1999;   99US-0139817.
PR   23-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   02-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.

```
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.


Query Match            46.9%;    Score 45;  DB 21;  Length 349;
Best Local Similarity  56.2%;    Pred. No. 28;
Matches    9;  Conservative  4;  Mismatches  4;  Indels  3;  Gaps  0;

Qy    4  LVLALAVALAVGDGSN 19
             |:: |:||:|||:|||
Db  169  llimaavllsvgegsn 184

Search completed: May  3, 2001, 15:52:21
Job time: 40 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.


OM protein - protein search, using sw model


Run on:      May  3, 2001, 15:51:41 ; Search time 11.61 Seconds
                                          (without alignments)
                                 34.748 Million cell updates/sec

Title:           US-09-426-776-10
Perfect score:   96
Sequence:        1 MRVLVLALAVALAVGDGSNLG 21

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters:       185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Issued_Patents_AA:*
              1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
              2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
              3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
              4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
              5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
              6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	45.8	629	1	US-08-250-740-33	Sequence 33, Appl
2	44	45.8	629	1	US-07-695-472B-2	Sequence 2, Appli
3	44	45.8	1027	4	US-08-446-137B-2	Sequence 2, Appli
4	43	44.8	222	2	US-08-391-743A-2	Sequence 2, Appli
5	43	44.8	683	1	US-07-878-960-2	Sequence 2, Appli
6	43	44.8	683	2	US-08-477-396A-17	Sequence 17, Appl
7	40	41.7	361	3	US-09-120-365-75	Sequence 75, Appl
8	40	41.7	516	2	US-08-676-166A-2	Sequence 2, Appli
9	40	41.7	777	2	US-08-477-396A-4	Sequence 4, Appli
10	40	41.7	1784	4	US-09-040-738-2	Sequence 2, Appli
11	39	40.6	455	1	US-08-738-172-4	Sequence 4, Appli
12	39	40.6	507	1	US-08-484-493-12	Sequence 12, Appl
13	39	40.6	507	1	US-08-484-494-12	Sequence 12, Appl
14	39	40.6	507	4	US-08-345-212-12	Sequence 12, Appl
15	39	40.6	507	4	US-09-249-003-12	Sequence 12, Appl
16	39	40.6	509	1	US-08-445-586-8	Sequence 8, Appli
17	38	39.6	90	1	US-09-124-671-25	Sequence 25, Appl
18	38	39.6	100	1	US-08-38-367-2	Sequence 2, Appli
19	38	39.6	109	4	US-09-124-671-23	Sequence 23, Appl
20	38	39.6	109	4	US-09-124-671-27	Sequence 27, Appl
21	38	39.6	109	4	US-09-124-671-29	Sequence 29, Appl
22	38	39.6	374	1	US-08-095-726-14	Sequence 14, Appl
23	38	39.6	374	4	US-08-096-623A-14	Sequence 14, Appl
24	38	39.6	1241	4	US-09-040-774-2	Sequence 2, Appli
25	37	38.5	214	1	US-07-953-230A-11	Sequence 11, Appl
26	37	38.5	402	3	US-08-948-997-4	Sequence 4, Appli
27	37	38.5	402	4	US-09-348-817A-4	Sequence 4, Appli
28	37	38.5	521	2	US-08-406-855A-19	Sequence 19, Appl
29	37	38.5	521	3	US-09-206-899-19	Sequence 19, Appl
30	37	38.5	559	2	US-08-406-855A-20	Sequence 20, Appl
31	37	38.5	559	3	US-09-206-899-20	Sequence 20, Appl
32	36.5	38.0	63	3	US-09-061-026-18	Sequence 18, Appl
33	36.5	38.0	63	4	US-09-466-138-18	Sequence 18, Appl
34	36	37.5	148	2	US-08-888-497-36	Sequence 36, Appl
35	36	37.5	148	5	PCT-US94-07926-36	Sequence 36, Appl
36	36	37.5	423	5	US-08-494-907-14	Sequence 14, Appl
37	36	37.5	423	5	PCT-US96-10986-14	Sequence 14, Appl
38	36	37.5	474	1	US-08-222-619-5	Sequence 5, Appli
39	36	37.5	474	5	PCT-US95-04075-5	Sequence 5, Appli
40	36	37.5	488	1	US-08-243-542-1	Sequence 1, Appli
41	36	37.5	488	1	US-08-477-407-1	Sequence 1, Appli
42	36	37.5	488	1	US-08-484-355-1	Sequence 1, Appli
43	36	37.5	524	1	US-08-243-542-2	Sequence 2, Appli
44	36	37.5	524	1	US-08-477-407-2	Sequence 2, Appli
45	36	37.5	524	1	US-08-484-355-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT  1
US-08-250-740-33
; Sequence 33, Application US/08250740
; Patent No. 5686240
; GENERAL INFORMATION:
;   APPLICANT: Schuchman, Edward H.
;   APPLICANT: Desnick, Robert J.
;   TITLE OF INVENTION: Acid Sphingomyelinase Gene and Diagnosis
;   TITLE OF INVENTION: of Niemann-Pick Disease
;   NUMBER OF SEQUENCES: 36
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Pennie & Edmonds
;     STREET: 1155 Avenue of the Americas
;     CITY: New York
;     STATE: New York
;     COUNTRY: USA
;     ZIP: 10036
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/250,740
;     FILING DATE: 27-MAY-1994
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Coruzzi, Laura A.
;     REGISTRATION NUMBER: 30742
;     REFERENCE/DOCKET NUMBER: 6923-038
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (212) 790-9090
;     TELEFAX: (212) 869-8864
;     TELEX: 66141 PENNIE
;   INFORMATION FOR SEQ ID NO: 33:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 629 amino acids
;       TYPE: amino acid
;       TOPOLOGY: unknown
;     MOLECULE TYPE: protein
US-08-250-740-33

Query Match             45.8%;  Score 44;  DB 1;  Length 629;
Best Local Similarity   71.4%;  Pred. No. 35;
Matches   10;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 VLVLALAVALAVGD 16
            || |||:|||:| |
Db     36 VLALALALALALSD 49
```

RESULT 2
US-07-695-472B-2
; Sequence 2, Application US/07695472B
; Patent No. 5773278
; GENERAL INFORMATION:
; APPLICANT: Schuchman, Edward H.
; APPLICANT: Desnick, Robert J.
; TITLE OF INVENTION: The Acid Sphingomyelinase Gene and
; TITLE OF INVENTION: Diagnosis of Niemann-Pick Disease
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/695,472B
; FILING DATE: 19910503
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6923-014
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 7908864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-07-695-472B-2

Query Match   45.8%; Score 44; DB 1; Length 629;
Best Local Similarity   71.4%; Pred. No. 35;
Matches   10; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

Qy   3 VLVLALAVALAVGD 16
        || |||:|||:| |
Db   36 VLALALALALALSD 49

RESULT 3
US-08-446-137B-2
; Sequence 2, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Trowern, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Duggleby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; TITLE OF INVENTION: FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,137B
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1027 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-446-137B-2

Query Match   45.8%; Score 44; DB 4; Length 1027;
Best Local Similarity   50.0%; Pred. No. 61;
Matches   9; Conservative   5; Mismatches   4; Indels   0; Gaps   0;

Qy   2 RVLVLALAVALAVGDGSN 19
        ::|:|||:| ||:| :|
Db   41 KLLMAALAGAIVVGGGAN 58

RESULT 4
US-08-391-743A-2
; Sequence 2, Application US/08391743A
; Patent No. 5843705
; GENERAL INFORMATION:
; APPLICANT: DiTullio, Paul A.; Meade, Harry; Cole, Edward S.
; TITLE OF INVENTION: TRANSGENETICALLY PRODUCED ANTITHROMBIN III
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/391,743A
; FILING DATE: 21-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: TCI-045
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid

```
Db        5 VRLLALALALGPAATL 22


RESULT  6
US-08-477-396A-17
; Sequence 17, Application US/08477396A
; Patent No. 5872235
; GENERAL INFORMATION:
;   APPLICANT:  Chen, Lan Bo
;   APPLICANT:  Bao, Shideng
;   APPLICANT:  Liu, Yuan
;   TITLE OF INVENTION:  A NOVEL TUMOR MARKER AND NOVEL METHOD OF
;   TITLE OF INVENTION:  ISOLATING SAME
;   NUMBER OF SEQUENCES:  19
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Weingarten, Schurgin, Gagnebin & Hayes
;     STREET:  Ten Post Office Square
;     CITY:  Boston
;     STATE:  Massachusetts
;     COUNTRY:  USA
;     ZIP:  02109
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0,  Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/477,396A
;     FILING DATE:
;   PRIOR APPLICATION DATA:
;     CLASSIFICATION:  435
;     APPLICATION NUMBER:  US 08/146,488
;     FILING DATE:  29-OCT-1993
;     APPLICATION NUMBER:  US 08/448,388
;     FILING DATE:  28-MAY-1996
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/US94/12502
;     FILING DATE:  31-OCT-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Heine, Holliday C.
;     REGISTRATION NUMBER:  34,346
;     REFERENCE/DOCKET NUMBER:  DFCI-333BX
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (617) 542-2290
;     TELEFAX:  (617) 451-0313
; INFORMATION FOR SEQ ID NO:  17:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  683 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
;   FRAGMENT TYPE:  internal
US-08-477-396A-17

Query Match            44.8%;   Score 43;  DB 2;  Length 683;
Best Local Similarity  55.6%;   Pred. No. 55;
Matches  10;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        3 VLVLALAVAVGDGSNL 20
            : :||||:|||:| : |
Db        5 VRLLALALALGPAATL 22


RESULT  7
US-09-120-365-75
; Sequence 75, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
;   APPLICANT:  Natori, Shunji
```

```
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-391-743A-2

Query Match            44.8%;   Score 43;  DB 2;  Length 222;
Best Local Similarity  64.3%;   Pred. No. 15;
Matches  9;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 MRVLVLALAVALAV 14
            |:||:|| ||||:
Db        1 MKVLILACLVALAI 14


RESULT  5
US-07-878-960-2
; Sequence 2, Application US/07878960
; Patent No. 5444164
; GENERAL INFORMATION:
;   APPLICANT:  Purchio, Anthony F.
;   APPLICANT:  Skonier, John
;   APPLICANT:  Neubauer, Michael G.
;   TITLE OF INVENTION:  TGF-BETA INDUCED GENE AND PROTEIN
;   NUMBER OF SEQUENCES:  2
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Bristol-Myers Squibb Company
;     STREET:  3005 First Avenue
;     CITY:  Seattle
;     STATE:  Washington
;     COUNTRY:  USA
;     ZIP:  98121
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0,  Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/878,960
;     FILING DATE:  05-MAY-1992
;     CLASSIFICATION:  530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/833,835
;     FILING DATE:  05-FEB-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Sorrentino, Joseph M.
;     REGISTRATION NUMBER:  32,598
;     REFERENCE/DOCKET NUMBER:  ON0092-
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  206/728-4800
;     TELEFAX:  206/727-3601
; INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  683 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
;   HYPOTHETICAL:  YES
;   FRAGMENT TYPE:  internal
;   ORIGINAL SOURCE:
;     ORGANISM:  Homo sapiens
;     TISSUE TYPE:  LUNG
;     CELL TYPE:  ADENOCARCINOMA
;     CELL LINE:  A549
US-07-878-960-2

Query Match            44.8%;   Score 43;  DB 1;  Length 683;
Best Local Similarity  55.6%;   Pred. No. 55;
Matches  10;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        3 VLVLALAVAVGDGSNL 20
            : :||||:|||:| : |
```

```
;   TITLE OF INVENTION: NEW PROTEASE
;   FILE REFERENCE: 32290-144749
;   CURRENT APPLICATION NUMBER: US/09/120,365
;   CURRENT FILING DATE: 1998-07-22
;   EARLIER APPLICATION NUMBER: JP 9-333 474
;   EARLIER FILING DATE: 1997-11-18
;   NUMBER OF SEQ ID NOS: 101
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
;   LENGTH: 361
;   TYPE: PRT
;   ORGANISM: Aleurain
US-09-120-365-75

Query Match       41.7%;    Score 40;   DB 3;   Length 361;
Best Local Similarity   54.5%;    Pred. No. 77;
Matches   12;   Conservative   3;   Mismatches   3;   Indels   4;   Gaps   1;

Qy      2  RVLVLALAV----ALAVGDGSN 19
           |||:|||||    |:|| |:
Db      5  RVLLALALVLATAAVAVASSSS 26

RESULT   8
US-08-676-166A-2
; Sequence 2, Application US/08676166A
; Patent No. 5955270
; GENERAL INFORMATION:
;   APPLICANT: Radford, Alan
;   APPLICANT: Parish, John H.
;   TITLE OF INVENTION: EXPLOITATION OF THE CELLULASE COMPLEX OF
;   TITLE OF INVENTION: NEUROSPORA
;   NUMBER OF SEQUENCES: 7
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: David A. Jackson, Esq.
;     STREET: 411 Hackensack Ave, Continental Plaza, 4th
;     STREET: Floor
;     CITY: Hackensack
;     STATE: New Jersey
;     COUNTRY: USA
;     ZIP: 07601
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/676,166A
;     FILING DATE: 15-JUL-1996
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Jackson Esq., David A.
;     REGISTRATION NUMBER: 26,742
;     REFERENCE/DOCKET NUMBER: 1321-1-002
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 201-487-5800
;     TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 516 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-676-166A-2

Query Match       41.7%;    Score 40;   DB 2;   Length 516;
Best Local Similarity   42.9%;    Pred. No. 1.2e+02;
Matches   9;   Conservative   3;   Mismatches   9;   Indels   0;   Gaps   0;

Qy      1  MRVLVLALAVALAVGDGSNLG 21
           ||  :|| ::|  || | | |
```

```
Db      1  MRASLLAFSLAAAVAGGQQAG 21

RESULT   9
US-08-477-396A-4
; Sequence 4, Application US/08477396A
; Patent No. 5872235
; GENERAL INFORMATION:
;   APPLICANT: Chen, Lan Bo
;   APPLICANT: Bao, Shideng
;   APPLICANT: Liu, Yuan
;   TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF
;   TITLE OF INVENTION: ISOLATING SAME
;   NUMBER OF SEQUENCES: 19
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
;     STREET: Ten Post Office Square
;     CITY: Boston
;     STATE: Massachusetts
;     COUNTRY: USA
;     ZIP: 02109
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/477,396A
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/146,488
;     FILING DATE: 29-OCT-1993
;     APPLICATION NUMBER: US 08/448,388
;     FILING DATE: 28-MAY-1996
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US94/12502
;     FILING DATE: 31-OCT-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Heine, Holliday C.
;     REGISTRATION NUMBER: 34,346
;     REFERENCE/DOCKET NUMBER: DFCI-333BX
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (617) 542-2290
;     TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 777 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-477-396A-4

Query Match       41.7%;    Score 40;   DB 2;   Length 777;
Best Local Similarity   56.2%;    Pred. No. 1.8e+02;
Matches   9;   Conservative   3;   Mismatches   4;   Indels   0;   Gaps   0;

Qy      5  VLALAVALAVGDGSNL 20
           :|||:|||:|
Db      7  ILALALALGPAVTL 22

RESULT   10
US-09-040-738-2
; Sequence 2, Application US/09040738
; Patent No. 6207374
; GENERAL INFORMATION:
;   APPLICANT: Sampson et al.
;   TITLE OF INVENTION: Tuberous Sclerosis 2 Gene and Uses
;   TITLE OF INVENTION: Thereof
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
```

ADDRESSEE: Kathleen M. Williams, Banner & Witcoff,
ADDRESSEE: Ltd.
STREET: One Financial Center
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,738
FILING DATE: Concurrently herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB9326470.3
FILING DATE: 24-December-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB9411900.5
FILING DATE: 14-June-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB94/02823
FILING DATE: -23-December-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/652,426
FILING DATE: 30-May-1996
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3265/73963
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-345-9100
TELEFAX: 617-345-9111
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1784 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-040-738-2

Query Match   41.7%; Score 40; DB 4; Length 1784;
Best Local Similarity 50.0%; Pred. No. 4.7e+02;
Matches   10; Conservative   1; Mismatches   9; Indels   0; Gaps   0;

Qy   2 RVLVLALAVAVGDGSNLG 21
         | |||| |: | || ||
Db 98 RHAVLALLKAIVGQGQGERLG 117

RESULT 11
US-08-738-172-4
; Sequence 4, Application US/08738172
; Patent No. 5939257
; GENERAL INFORMATION:
APPLICANT: Szasz, Joseph
APPLICANT: Davis, Maria
TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage

COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738,172
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/005,965
FILING DATE: October 27, 1995
APPLICATION NUMBER: 08/465,003
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/575,354
FILING DATE: December 20, 1995
APPLICATION NUMBER: 08/240,158
FILING DATE: May 10, 1994
APPLICATION NUMBER: 08/229,329
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/158
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-738-172-4

Query Match   40.6%; Score 39; DB 2; Length 455;
Best Local Similarity 44.4%; Pred. No. 1.4e+02;
Matches   8; Conservative   5; Mismatches   5; Indels   0; Gaps   0;

Qy   1 MRVLVLALAVALAVGDGS 18
         || ::|| : | ||::|:
Db   1 MRWILLFLVMLLGVGEGA 18

RESULT 12
US-08-484-493-12
; Sequence 12, Application US/08484493
; Patent No. 5728381
; GENERAL INFORMATION:
APPLICANT: Wilson, Peter J
APPLICANT: Morris, Charles P
APPLICANT: Anson, Donald S
APPLICANT: Occhiodoro, Teresa
APPLICANT: Bielicki, Julie
APPLICANT: Clements, Peter R
APPLICANT: Hopwood, John J
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
TITLE OF INVENTION: IDURONATE 2-SULFATASE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,493
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 991,973
FILING DATE: 17-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: DiGiglio, Frank S
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8416Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-493-12

Query Match   40.6%;   Score 39; DB 1; Length 507;
Best Local Similarity   52.6%;   Pred. No. 1.6e+02;
Matches   10; Conservative   2; Mismatches   7; Indels   0; Gaps   0;

Qy   2 RVLVLALAVALAVGDGSNL 20
         | |:|||| ||| |:
Db   5 RSLLLALAAGLAVARPPNI 23

RESULT 13
US-08-484-494-12
; Sequence 12, Application US/08484494
; Patent No. 5798239
; GENERAL INFORMATION:
; APPLICANT: Wilson, Peter J
; APPLICANT: Morris, Charles P
; APPLICANT: Anson, Donald S
; APPLICANT: Occhiodoro, Teresa
; APPLICANT: Bielicki, Julie
; APPLICANT: Clements, Peter R
; APPLICANT: Hopwood, John J
; TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
; TITLE OF INVENTION: IDURONATE 2-SULFATASE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,494
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 991,973
; FILING DATE: 17-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DiGiglio, Frank S
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8416Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343

TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-494-12

Query Match   40.6%;   Score 39; DB 1; Length 507;
Best Local Similarity   52.6%;   Pred. No. 1.6e+02;
Matches   10; Conservative   2; Mismatches   7; Indels   0; Gaps   0;

Qy   2 RVLVLALAVALAVGDGSNL 20
         | |:|||| ||| |:
Db   5 RSLLLALAAGLAVARPPNI 23

RESULT 14
US-08-345-212-12
; Sequence 12, Application US/08345212
; Patent No. 5932211
; GENERAL INFORMATION:
; APPLICANT: Wilson, Peter J
; APPLICANT: Morris, Charles P
; APPLICANT: Anson, Donald S
; APPLICANT: Occhiodoro, Teresa
; APPLICANT: Bielicki, Julie
; APPLICANT: Clements, Peter R
; APPLICANT: Hopwood, John J
; TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
; TITLE OF INVENTION: IDURONATE 2-SULFATASE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/345,212
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 991,973
; FILING DATE: 17-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DiGiglio, Frank S
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8416Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-345-212-12

Query Match   40.6%; Score 39; DB 2; Length 507;
Best Local Similarity 52.6%; Pred. No. 1.6e+02;
Matches   10; Conservative   2; Mismatches   7; Indels   0; Gaps   0;

Qy   2 RVLVLALAVALAVGDGSNL 20
        | |:|||| ||| |:
Db   5 RSLLLALAAGLAVARPPNI 23

RESULT 15
US-09-249-003-12
; Sequence 12, Application US/09249003
; Patent No. 6153188
; GENERAL INFORMATION:
;   APPLICANT: Wilson, Peter J
;   APPLICANT: Morris, Charles P
;   APPLICANT: Anson, Donald S
;   APPLICANT: Occhiodoro, Teresa
;   APPLICANT: Bielicki, Julie
;   APPLICANT: Clements, Peter R
;   APPLICANT: Hopwood, John J
;   TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
;   TITLE OF INVENTION: IDURONATE 2-SULFATASE
;   NUMBER OF SEQUENCES: 15
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Scully, Scott, Murphy & Presser
;   STREET: 400 Garden City Plaza
;   CITY: Garden City
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 11530
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/249,003
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/991,973
;   FILING DATE: 17-DEC-1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME: DiGiglio, Frank S
;   REGISTRATION NUMBER: 31,346
;   REFERENCE/DOCKET NUMBER: 8416Z
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 516-742-4343
;   TELEFAX: 516-742-4366
;   TELEX: 230 901 SANS UR
;   INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 507 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-249-003-12

Query Match   40.6%; Score 39; DB 4; Length 507;
Best Local Similarity 52.6%; Pred. No. 1.6e+02;
Matches   10; Conservative   2; Mismatches   7; Indels   0; Gaps   0;

Qy   2 RVLVLALAVALAVGDGSNL 20
        | |:|||| ||| |:
Db   5 RSLLLALAAGLAVARPPNI 23

Search completed: May 3, 2001, 15:51:59
Job time: 18 sec

```
                    GenCore version 4.5
              Copyright (c) 1993 - .2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        May  5, 2001, 21:23:16 ; Search time 1155.53 Seconds
                                        (without alignments)
                                        604.820 Million cell updates/sec

Title:         US-09-426-776-11
Perfect score: 80
Sequence:      1 attcacatcccacagccatg..........gggaccagtccaacttgggg 80

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters:      19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    EST:*
               1: gb_est1:*
               2: gb_est2:*
               3: gb_est3:*
               4: gb_est4:*
               5: gb_est5:*
               6: gb_est6:*
               7: gb_est7:*
               8: gb_est8:*
               9: gb_est9:*
              10: gb_est10:*
              11: gb_est11:*
              12: gb_est12:*
              13: gb_est13:*
              14: gb_est14:*
              15: gb_est15:*
              16: gb_est16:*
              17: gb_est17:*
              18: gb_est18:*
              19: gb_est19:*
              20: gb_est20:*
              21: gb_est21:*
              22: gb_est22:*
              23: gb_est23:*
              24: gb_est24:*
              25: gb_est33:*
              26: gb_est34:*
              27: gb_est35:*
              28: gb_est36:*
              29: gb_est37:*
              30: gb_est38:*
              31: gb_est39:*
              32: gb_est40:*
              33: em_estba:*
              34: em_estfun:*
              35: em_esthum1:*
              36: em_esthum2:*
              37: em_esthum3:*
              38: em_esthum4:*
              39: em_esthum5:*
              40: em_esthum6:*
              41: em_esthum7:*
              42: em_esthum8:*
              43: em_esthum9:*
              44: em_esthum10:*
              45: em_esthum11:*
              46: em_esthum12:*
              47: em_esthum13:*
              48: em_esthum14:*
              49: em_esthum15:*
              50: em_esthum16:*
              51: em_esthum17:*
              52: em_esthum18:*
              53: em_esthum19:*
              54: em_esthum20:*
              55: em_esthum21:*
              56: em_esthum22:*
              57: em_esthum23:*
              58: em_esthum24:*
              59: em_esthum25:*
              60: em_esthum26:*
              61: em_esthum27:*
              62: em_esthum28:*
              63: em_estin1:*
              64: em_estin2:*
              65: em_estin3:*
              66: em_estin4:*
              67: em_estin5:*
              68: em_estom1:*
              69: em_estom2:*
              70: em_estov1:*
              71: em_estov2:*
              72: em_estpl1:*
              73: em_estpl2:*
              74: em_estpl3:*
              75: em_estpl4:*
              76: em_estpl5:*
              77: em_estpl6:*
              78: em_estpl7:*
              79: em_estpl8:*
              80: em_estpl9:*
              81: em_estpl10:*
              82: em_estro1:*
              83: em_estro2:*
              84: em_estro3:*
              85: em_estro4:*
              86: em_estro5:*
              87: em_estro6:*
              88: em_estro7:*
              89: em_estro8:*
              90: em_estro9:*
              91: em_estro10:*
              92: em_estro11:*
              93: em_estro12:*
              94: em_estro13:*
              95: em_estro14:*
              96: em_estro15:*
              97: em_estro16:*
              98: em_estro17:*
              99: em_estro18:*
             100: em_estro19:*
             101: em_estro20:*
             102: gb_est25:*
             103: gb_est26:*
             104: gb_est27:*
             105: gb_est28:*
             106: gb_est29:*
             107: gb_est30:*
             108: gb_est31:*
             109: gb_est32:*
             110: gb_est41:*
             111: gb_est42:*
             112: gb_est43:*
             113: gb_est44:*
             114: gb_est45:*
             115: gb_est46:*
             116: gb_est47:*
```

```
117:  gb_est48:*            190:  em_gss_pln1:*
118:  gb_est49:*            191:  em_gss_pln2:*
119:  gb_est50:*            192:  em_gss_pro:*
120:  gb_est51:*            193:  em_gss_rod1:*
121:  gb_est52:*            194:  em_gss_rod2:*
122:  gb_est53:*            195:  em_gss_rod3:*
123:  gb_est54:*            196:  em_gss_rod4:*
124:  gb_est55:*            197:  em_gss_rod5:*
125:  gb_est56:*            198:  em_gss_vrt1:*
126:  gb_est57:*            199:  em_gss_vrt2:*
127:  gb_est58:*            200:  em_gss_vrt3:*
128:  gb_est59:*            201:  gb_gss1:*
129:  gb_est60:*            202:  gb_gss2:*
130:  gb_est61:*            203:  gb_gss3:*
131:  gb_est62:*            204:  gb_gss4:*
132:  gb_est63:*            205:  gb_gss5:*
133:  gb_est64:*            206:  gb_gss6:*
134:  gb_est65:*            207:  gb_gss7:*
135:  gb_est66:*            208:  gb_gss8:*
136:  gb_est75:*            209:  gb_gss9:*
137:  gb_est76:*            210:  gb_gss10:*
138:  gb_est77:*            211:  gb_gss11:*
139:  gb_est78:*            212:  gb_gss12:*
140:  gb_est79:*            213:  gb_gss13:*
141:  gb_est80:*            214:  gb_gss14:*
142:  gb_est81:*            215:  gb_gss15:*
143:  gb_est82:*            216:  gb_gss16:*
144:  gb_est83:*            217:  gb_gss17:*
145:  gb_est84:*            218:  gb_gss18:*
146:  gb_est85:*            219:  gb_gss19:*
147:  gb_est86:*            220:  gb_gss20:*
148:  gb_est87:*            221:  gb_gss21:*
149:  gb_est88:*            222:  gb_gss22:*
150:  gb_est89:*            223:  gb_gss23:*
151:  gb_est90:*            224:  gb_gss24:*
152:  gb_est99:*            225:  gb_gss25:*
153:  gb_est100:*           226:  gb_gss26:*
154:  gb_est101:*           227:  gb_gss27:*
155:  gb_est102:*           228:  gb_gss28:*
156:  gb_est103:*           229:  gb_gss29:*
157:  gb_est104:*           230:  gb_gss30:*
158:  gb_est105:*           231:  gb_gss31:*
159:  gb_est106:*           232:  gb_gss32:*
160:  gb_est67:*            233:  gb_gss33:*
161:  gb_est68:*            234:  gb_gss34:*
162:  gb_est69:*
163:  gb_est70:*
164:  gb_est71:*
165:  gb_est72:*
166:  gb_est73:*
167:  gb_est74:*
168:  gb_est91:*
169:  gb_est92:*
170:  gb_est93:*
171:  gb_est94:*
172:  gb_est95:*
173:  gb_est96:*
174:  gb_est97:*
175:  gb_est98:*
176:  em_gss_fun:*
177:  em_gss_hum1:*
178:  em_gss_hum2:*
179:  em_gss_hum3:*
180:  em_gss_hum4:*
181:  em_gss_hum5:*
182:  em_gss_hum6:*
183:  em_gss_hum7:*
184:  em_gss_hum8:*
185:  em_gss_hum9:*
186:  em_gss_inv1:*
187:  em_gss_inv2:*
188:  em_gss_inv3:*
189:  em_gss_other:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	
1	33.6	42.0	489	144	BF156722	BF156722	f156a04.y
2	33.6	42.0	502	144	BF157180	BF157180	f135g04.y
3	33.6	42.0	559	111	AW133798	AW133798	f112a02.y
4	33.6	42.0	633	144	BF156786	BF156786	f156g07.y
5	33.6	42.0	643	144	BF156818	BF156818	f157c10.y
6	33.6	42.0	718	144	BF157814	BF157814	f137a10.y
7	33.6	42.0	727	144	BF156997	BF156997	f159e07.y
8	33.6	42.0	727	144	BF158524	BF158524	f139g04.y
9	33.6	42.0	736	144	BF156434	BF156434	f152a06.y
10	33.6	42.0	763	144	AW153678	AW153678	f124e02.y
11	33.6	42.0	763	111	AW175005	AW175005	f131b10.y
12	33.6	42.0	764	111	AW153397	AW153397	f120c02.y
13	33.4	41.8	614	111	AW128094	AW128094	f106f12.y
14	32.6	40.8	740	144	BF157419	BF157419	f138b06.y
15	32.4	40.5	667	114	AW343221	AW343221	f174g10.y
c 16	32.2	40.3	796	226	A2697043	A2697043	RPCI-23-2
17	32	40.0	304	144	BF156771	BF156771	f156f01.y
18	32	40.0	415	114	AW343413	AW343413	f177c08.y

19	32	40.0	438	144	BF156728	BF156728 f156a11.y
20	32	40.0	459	144	BF157193	BF157193 f135h06.y
21	32	40.0	469	144	BF156380	BF156380 f151d01.y
22	32	40.0	484	144	BF158075	BF158075 f128b12.y
23	32	40.0	503	144	BF156530	BF156530 f153f02.y
24	32	40.0	511	144	BF158308	BF158308 f131b12.y
25	32	40.0	544	144	BF156971	BF156971 f159c01.y
26	32	40.0	553	144	BF156196	BF156196 f161a04.y
27	32	40.0	560	111	AW133980	AW133980 f114a12.y
28	32	40.0	592	144	BF158418	BF158418 f132e08.y
29	32	40.0	594	111	AW133689	AW133689 fi10e11.y
30	32	40.0	598	24	AI722091	AI722091 fd19b12.y
31	32	40.0	598	111	AW175410	AW175410 f135g04.y
32	32	40.0	601	144	BF157832	BF157832 f137c08.y
33	32	40.0	604	111	AW133909	AW133909 fi13c08.y
34	32	40.0	604	144	BF158373	BF158373 f132a04.y
35	32	40.0	607	24	AI722220	AI722220 fd20e12.y
36	32	40.0	607	144	BF156221	BF156221 f161c07.y
37	32	40.0	607	144	BF156230	BF156230 f161d05.y
38	32	40.0	607	144	BF156413	BF156413 f151g04.y
39	32	40.0	607	144	BF157066	BF157066 f160d01.y
40	32	40.0	607	144	BF157517	BF157517 f140e08.y
41	32	40.0	608	144	BF156292	BF156292 f150b11.y
42	32	40.0	610	24	AI721781	AI721781 fd16c11.y
43	32	40.0	613	111	AW174217	AW174217 f140e10.y
44	32	40.0	613	144	BF157445	BF157445 f138e10.y
45	32	40.0	613	144	BF157975	BF157975 f127a05.y

ALIGNMENTS

RESULT 1
BF156722
LOCUS   BF156722   489 bp   mRNA   EST   30-OCT-2000
DEFINITION f156a04.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            3817638 5' similar to TR:093605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   BF156722
VERSION   BF156722.1 GI:11051909
KEYWORDS   EST.
SOURCE   zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 489)
AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
TITLE   WashU Zebrafish EST Project 1998
JOURNAL   Unpublished (1998)
COMMENT   Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution information can be found through the I.M.A.G.E.
            Consortium/LLNL, send email to: info@image.llnl.gov
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 487.
FEATURES   Location/Qualifiers
     source   1..489
                     /organism="Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="3817638"

                     /clone_lib="Sugano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, including
                     unfertilized eggs)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fragments
                     <1.5kb. Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Koichi Kawakami. Custom primers for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."

BASE COUNT   117 a   134 c   116 g   121 t   1 others
ORIGIN

Query Match   42.0%;   Score 33.6;   DB 144;   Length 489;
Best Local Similarity   66.7%;   Pred. No. 0.54;
Matches   48;   Conservative   0;   Mismatches   24;   Indels   0;   Gaps   0;

Qy   7 atccaccagccatgaggtgcttgtactagctccttcgtggctctcgcagtggggacc 66
          | ||||||||||||||||| | ||||| ||| || || | |||| | || ||| |
Db   5 AACCACCAGCCATGAGAGCTGTTGCCCTGACTGTGCGCTGAGCCCTCGTGGCGAGTCAAC 64

Qy   67 agtccaacttgg 78
          || ||| | |
Db   65 AGATGAACCTTG 76

RESULT 2
BF157180
LOCUS   BF157180   502 bp   mRNA   EST   30-OCT-2000
DEFINITION f135g04.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            3815983 5' similar to TR:093605 093605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   BF157180
VERSION   BF157180.1 GI:11052370
KEYWORDS   EST.
SOURCE   zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 502)
AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
TITLE   WashU Zebrafish EST Project 1998
JOURNAL   Unpublished (1998)
COMMENT   Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution information can be found through the I.M.A.G.E.
            Consortium/LLNL, send email to: info@image.llnl.gov
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 489.
FEATURES   Location/Qualifiers
     source   1..502
                     /organism="Danio rerio"

```
                    /strain="AB"
                    /db_xref="taxon:7955"
                    /clone="3815983"
                    /clone_lib="Sugano Kawakami zebrafish DRA"
                    /sex="mixed (one male and one female, including
                    unfertilized eggs)"
                    /dev_stage="adult"
                    /lab_host="DH10B (phage resistant)"
                    /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                    Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                    with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTT];
                    double-stranded cDNA was ligated to a DraIII adaptor
                    [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                    sites of the pME18S-FL3 vector (5' site CACCATGTG, 3' site
                    CACCATGTG). XhoI should be used to isolate the cDNA
                    insert. Size selection was performed to exclude fragments
                    <1.5kb. Library constructed by Dr. Sumio Sugano
                    (University of Tokyo Institute of Medical Science) and
                    kindly donated by Dr. Koichi Kawakami. Custom primers for
                    sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                    primer CGACCTGCAGCTCGAGCACA. "

BASE COUNT      125 a    129 c    119 g    129 t
ORIGIN

Query Match            42.0%;    Score 33.6; DB 144; Length 502;
Best Local Similarity  66.7%;    Pred. No. 0.55;
Matches      48;  Conservative   0;  Mismatches  24;  Indels   0;  Gaps   0;

Qy      7 atccaccagccatgagggtgcttgtactagctcttgctgtggctctcgcagtgggggacc 66
          | |||||||||||||||| |||| || ||  |   |||| || |||| |
Db      3 AACCACCAGCCATGAGAGCGTGTTGTGCTTGCCCTGACTGTAGCCCTCGTGGCCATGTCAAC 62

Qy     67 agtccaacttgg 78
          || ||||||| |
Db     63 AATTCAACCTTG 74

RESULT 3
AW133798
LOCUS       AW133798      559 bp   mRNA        EST      27-OCT-1999
DEFINITION  fi12a02.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            2600906 5' similar to TR:O93605 O93605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   AW133798
VERSION     AW133798.1  GI:6135405
KEYWORDS    EST.
SOURCE      zebrafish.
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 559)
AUTHORS     Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter
            ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE       WashU Zebrafish EST Project 1999
JOURNAL     Unpublished (1999)
COMMENT     Contact: S.L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 527.
FEATURES    Location/Qualifiers
source      1..559
            /organism="Danio rerio"
            /strain="AB"
```

```
                    /db_xref="taxon:7955"
                    /clone="2600906"
                    /clone_lib="Sugano Kawakami zebrafish DRA"
                    /sex="mixed (one male and one female, including
                    unfertilized eggs)"
                    /dev_stage="adult"
                    /lab_host="DH10B (phage resistant)"
                    /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                    Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                    with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTT];
                    double-stranded cDNA was ligated to a DraIII adaptor
                    [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                    sites of the pME18S-FL3 vector (5' site CACCATGTG, 3' site
                    CACCATGTG). XhoI should be used to isolate the cDNA
                    insert. Size selection was performed to exclude fragments
                    <1.5kb. Library constructed by Dr. Sumio Sugano
                    (University of Tokyo Institute of Medical Science) and
                    kindly donated by Dr. Koichi Kawakami. Custom primers for
                    sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                    primer CGACCTGCAGCTCGAGCACA. "

BASE COUNT      144 a    151 c    128 g    136 t
ORIGIN

Query Match            42.0%;    Score 33.6; DB 111; Length 559;
Best Local Similarity  66.7%;    Pred. No. 0.56;
Matches      48;  Conservative   0;  Mismatches  24;  Indels   0;  Gaps   0;

Qy      7 atccaccagccatgagggtgcttgtactagctcttgctgtggctctcgcagtgggggacc 66
          | |||||||||||||||| |||| || ||  |   |||| || |||| |
Db      4 AACCACCAGCCATGAGAGCGTGTTGTGCTTGCCCTGACTGTAGCCCTCGTGGCCAGTCAAC 63

Qy     67 agtccaacttgg 78
          || ||||||| |
Db     64 AGATGAACCTTG 75

RESULT 4
BF156786
LOCUS       BF156786      633 bp   mRNA        EST      30-OCT-2000
DEFINITION  fi56g07.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            3817932 5' similar to TR:O93605 O93605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   BF156786
VERSION     BF156786.1  GI:11051974
KEYWORDS    EST.
SOURCE      zebrafish.
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 633)
AUTHORS     Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
TITLE       WashU Zebrafish EST Project 1998
JOURNAL     Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            distribution information can be found through the I.M.A.G.E.
            Consortium/LLNL, send email to: info@image.llnl.gov
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 496.
FEATURES    Location/Qualifiers
```

```
source          1..633
                /organism="Danio rerio"
                /strain="AB"
                /db_xref="taxon:7955"
                /clone="3817932"
                /clone_lib="Sugano Kawakami zebrafish DRA"
                /sex="mixed (one male and one female, including
                unfertilized eggs)"
                /dev_stage="adult"
                /lab_host="DH10B (phage resistant)"
                /note="vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT];
                double-stranded cDNA was ligated to a DraIII adaptor
                [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                CACCATGTG). XhoI should be used to isolate the cDNA
                insert. Size selection was performed to exclude fragments
                <1.5kb. Library constructed by Dr. Sumio Sugano
                (University of Tokyo Institute of Medical Science) and
                kindly donated by Dr. Koichi Kawakami. Custom primers for
                sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                primer CGACCTGCAGCTCGAGCACA. "
BASE COUNT      189 a      159 c     141 g    144 t
ORIGIN

Query Match               42.0%;     Score 33.6;  DB 144;  Length 633;
Best Local Similarity     66.7%;     Pred. No. 0.57;
Matches    48;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      7 atccaccagccatgagggtgcttgtactagcttcttgctggctctcgcagtgggggacc 66
          | ||||||||||||||| |   |||| || ||| || ||||| |||| ||||  | ||
Db     34 AACCACCAGCCATGAGAGCTGTTGTGCCTGCCTGACTGTAGCCCCTGCGTGGCGAGTCAAC 93

Qy     67 agtccaacttgg 78
          || ||| ||
Db     94 AGATGAACCTTG 105

RESULT 5
BF156818
LOCUS       BF156818      643 bp      mRNA      EST      30-OCT-2000
DEFINITION  f157c10.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            3817747 5' similar to TR:O93605 O93605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   BF156818
VERSION     BF156818.1   GI:11052006
KEYWORDS    EST.
SOURCE      zebrafish.
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 643)
AUTHORS     Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
TITLE       WashU zebrafish EST Project 1998
JOURNAL     Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution information can be found through the I.M.A.G.E.
            Consortium/LLNL, send email to: info@image.llnl.gov
```

```
                Seq primer: T3 ET from Amersham
                High quality sequence stop: 531.
FEATURES             Location/Qualifiers
     source          1..643
                     /organism="Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="3817747"
                     /clone_lib="Sugano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, including
                     unfertilized eggs)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /note="vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fragments
                     <1.5kb. Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Koichi Kawakami. Custom primers for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA. "
BASE COUNT      169 a      167 c     154 g    153 t
ORIGIN

Query Match               42.0%;     Score 33.6;  DB 144;  Length 643;
Best Local Similarity     66.7%;     Pred. No. 0.58;
Matches    48;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      7 atccaccagccatgagggtgcttgtactagcttcttgctggctctcgcagtgggggacc 66
          | ||||||||||||||| |   |||| || ||| || ||||| |||| ||||  | ||
Db      3 AACCACCAGCCATGAGAGCTGTTGTGCCTGCCTGACTGTAGCCCCTGCGTGGCGAGTCAAC 62

Qy     67 agtccaacttgg 78
          || ||| ||
Db     63 AGATGAACCTTG 74

RESULT 6
BF157814
LOCUS       BF157814      718 bp      mRNA      EST      30-OCT-2000
DEFINITION  f137a10.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            3815731 5' similar to TR:O93605 O93605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   BF157814
VERSION     BF157814.1   GI:11053013
KEYWORDS    EST.
SOURCE      zebrafish.
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 718)
AUTHORS     Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
TITLE       WashU zebrafish EST Project 1998
JOURNAL     Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
```

Sequencing by: Washington University Genome Sequencing Center Clone
distribution information can be found through the I.M.A.G.E.
Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 485.
FEATURES   Location/Qualifiers
source   1..718
             /organism="Danio rerio"
             /strain="AB"
             /db_xref="taxon:7955"
             /clone="3815731"
             /clone_lib="Sugano Kawakami zebrafish DRA"
             /sex="mixed (one male and one female, including
             unfertilized eggs)"
             /dev_stage="adult"
             /lab_host="DH10B (phage resistant)"
             /note="vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
             Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
             with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTT];
             double-stranded cDNA was ligated to a DraIII adaptor
             [TGTTGGCCTACTCG], digested and cloned into distinct DraIII
             sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
             CACCATGTG). XhoI should be used to isolate the cDNA
             insert. Size selection was performed to exclude fragments
             <1.5kb. Library constructed by Dr. Sumio Sugano
             (University of Tokyo Institute of Medical Science) and
             kindly donated by Dr. Koichi Kawakami. Custom primers for
             sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
             primer CGACCTGCAGCTCGAGCACA."
BASE COUNT   194 a   183 c   169 g   171 t   1 others
ORIGIN

Query Match   42.0%;   Score 33.6; DB 144; Length 718;
Best Local Similarity   66.7%;   Pred. No. 0.59;
Matches   48; Conservative   0; Mismatches   24; Indels   0; Gaps   0;

Qy   7 atccaccagccatgagggtgctgtactagctcttgctgtggctctcgcagtgggggacc 66
          | |||||||||||||| | |||| ||| | || ||| |||| || | |||| || |
Db   3 AACCACCAGCCATGAGAGCTGTTGTGCTTGCCCTGACTGTAGCCCTCGTGGGCGAGTCAAC 62

Qy   67 agtccaacttgg 78
          || ||| |
Db   63 AGATGAACCTTG 74

RESULT 7
BF156997
LOCUS   BF156997   727 bp   mRNA   EST   30-OCT-2000
DEFINITION   f159e07.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
             3818197 5' similar to TR:O93605 O93605 VITELLOGENIN ;, mRNA
             sequence.
ACCESSION   BF156997
VERSION   BF156997.1 GI:11052186
KEYWORDS   EST.
SOURCE   zebrafish.
ORGANISM   Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
             Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 727)
AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
             ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
             ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
             Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
             Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
             and Wilson,R.
TITLE   WashU Zebrafish EST Project 1998
JOURNAL   Unpublished (1998)
COMMENT   Contact: Stephen L. Johnson
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
             Tel: 314 286 1800

TITLE   WashU zebrafish EST Project 1998
JOURNAL   Unpublished (1998)
COMMENT   Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution information can be found through the I.M.A.G.E.
            Consortium/LLNL, send email to: info@image.llnl.gov
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 518.
FEATURES
     source   1..736
                     /organism="Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="3817258"
                     /clone_lib="Sugano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, including
                     unfertilized eggs)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fragments
                     <1.5kb. Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Koichi Kawakami. Custom primers for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT   199 a   185 c   178 g   174 t
ORIGIN

Query Match   42.0%; Score 33.6; DB 144; Length 736;
Best Local Similarity 66.7%; Pred. No. 0.59;
Matches   48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy   7 atccaccagccatgagggtgcttgtactagctcttgtctggcctctcgcagtgggggacc 66
         | |||||||||||||| | |||| || |||| || |||| || |||| | | |
Db   3 AACCACCAGCCATGAGAGAGCGTGTTGTGTTCGCCCTGACTGTAGCCCTCCGTGGCGAGTCAAC 62

Qy   67 agtccaacttgg 78
         || ||| |
Db   63 AGATGAACCTTG 74

RESULT 10
AW153678
LOCUS   AW153678   763 bp   mRNA   EST   03-NOV-1999
DEFINITION f124e02.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            2602106 5' similar to TR:O93605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   AW153678
VERSION   AW153678.1 GI:6201578
KEYWORDS   EST.
SOURCE   zebrafish.
  ORGANISM Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 763)
  AUTHORS   Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter

---

            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution information can be found through the I.M.A.G.E.
            Consortium/LLNL, send email to: info@image.llnl.gov
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 508.
FEATURES
     source   1..727
                     /organism="Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="381563"
                     /clone_lib="Sugano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, including
                     unfertilized eggs)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a DraIII adaptor.
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fragments
                     <1.5kb. Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Koichi Kawakami. Custom primers for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA.   3 others
BASE COUNT   196 a   184 c   170 g   174 t
ORIGIN

Query Match   42.0%; Score 33.6; DB 144; Length 727;
Best Local Similarity 66.7%; Pred. No. 0.59;
Matches   48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy   7 atccaccagccatgagggtgcttgtactagctcttgtctggcctctcgcagtgggggacc 66
         | |||||||||||||| | |||| || |||| || |||| || |||| | | |
Db   3 AACCACCAGCCATGAGAGAGCGTGTTGTGTTCGCCCTGACTGTAGCCCTCCGTGGCGAGTCAAC 62

Qy   67 agtccaacttgg 78
         || ||| |
Db   63 AGATGAACCTTG 74

RESULT 9
BF156434
LOCUS   BF156434   736 bp   mRNA   EST   30-OCT-2000
DEFINITION f152a06.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            3817258 5' similar to TR:O93605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   BF156434
VERSION   BF156434.1 GI:11051618
KEYWORDS   EST.
SOURCE   zebrafish.
  ORGANISM Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 736)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,V., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.

Column 1 (RESULT 11)

```
              ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE       WashU Zebrafish EST Project 1999
JOURNAL     Unpublished (1999)
COMMENT     Contact: S.L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 524.
FEATURES    Location/Qualifiers
   source   1..763
            /organism="Danio rerio"
            /strain="AB"
            /db_xref="taxon:7955"
            /clone="2602106"
            /clone_lib="Sugano Kawakami zebrafish DRA"
            /sex="mixed (one male and one female, including
            unfertilized eggs)"
            /dev_stage="adult"
            /lab_host="DH10B (phage resistant)"
            /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
            Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
            with an oligo(dT) primer [ATGTGGCCGTTTTTTTTTTTTTTTT];
            double-stranded cDNA was ligated to a DraIII adaptor
            [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
            sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
            CACCATGTG). XhoI should be used to isolate the cDNA
            insert.  Size selection was performed to exclude fragments
            <1.5kb.  Library constructed by Dr. Sumio Sugano
            (University of Tokyo Institute of Medical Science) and
            kindly donated by Dr. Koichi Kawakami.  Custom primers for
            sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end
            primer CGACCTGCAGCTCGAGCACA. "
BASE COUNT    206 a    193 c    179 g    184 t    1 others
ORIGIN

Query Match             42.0%;  Score 33.6;  DB 111;  Length 763;
Best Local Similarity   66.7%;  Pred. No. 0.6;
Matches  48;  Conservative  0;  Mismatches  24;  Indels  0;  Gaps  0;

Qy      7 atccacagccatgagggtgctgtactagctcttgctgtggcctctgcagtggggacc 66
          | |||||||||||||| ||| || ||||| || ||| |||| |
Db      3 AACCACCAGCCATGAGAGAGCTGTTGTGCTTGCCCCTGACTGTAGCCCTCGTGGCCAGTCAAC 62

Qy     67 agtccaacttgg 78
          || ||| |
Db     63 AGATGAACCTTG 74


RESULT 11
AW175005
LOCUS       AW175005        763 bp    mRNA          EST        16-NOV-1999
DEFINITION  fi3lb10.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            2639227 5' similar to TR:O93605 O93605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   AW175005
VERSION     AW175005.1  GI:6441025
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 763)
  AUTHORS   Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter
            ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
```

Column 2 (RESULT 12)

```
TITLE       WashU Zebrafish EST Project 1999
JOURNAL     Unpublished (1999)
COMMENT     Other_ESTs: fi3lb10.x1
            Contact: S.L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 520.
FEATURES    Location/Qualifiers
   source   1..763
            /organism="Danio rerio"
            /strain="AB"
            /db_xref="taxon:7955"
            /clone="2639227"
            /clone_lib="Sugano Kawakami zebrafish DRA"
            /sex="mixed (one male and one female, including
            unfertilized eggs)"
            /dev_stage="adult"
            /lab_host="DH10B (phage resistant)"
            /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
            Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
            with an oligo(dT) primer [ATGTGGCCGTTTTTTTTTTTTTTTT];
            double-stranded cDNA was ligated to a DraIII adaptor
            [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
            sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
            CACCATGTG). XhoI should be used to isolate the cDNA
            insert.  Size selection was performed to exclude fragments
            <1.5kb.  Library constructed by Dr. Sumio Sugano
            (University of Tokyo Institute of Medical Science) and
            kindly donated by Dr. Koichi Kawakami.  Custom primers for
            sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end
            primer CGACCTGCAGCTCGAGCACA. "
BASE COUNT    207 a    191 c    182 g    183 t
ORIGIN

Query Match             42.0%;  Score 33.6;  DB 111;  Length 763;
Best Local Similarity   66.7%;  Pred. No. 0.6;
Matches  48;  Conservative  0;  Mismatches  24;  Indels  0;  Gaps  0;

Qy      7 atccacagccatgagggtgctgtactagctcttgctgtggcctctcgcagtggggacc 66
          | |||||||||||||| ||| || ||||| || ||| |||| |
Db      3 AACCACCAGCCATGAGAGAGCTGTTGTGCTTGCCCCTGACTGTAGCCCTCGTGGCCAGTCAAC 62

Qy     67 agtccaacttgg 78
          || ||| |
Db     63 AGATGAACCTTG 74


RESULT 12
AW153397
LOCUS       AW153397        764 bp    mRNA          EST        03-NOV-1999
DEFINITION  fi120c02.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            2601698 5' similar to TR:O93605 O93605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   AW153397
VERSION     AW153397.1  GI:6201297
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 764)
  AUTHORS   Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter
            ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
```

TITLE   WashU Zebrafish EST Project 1999
JOURNAL   Unpublished (1999)
COMMENT   Other_ESTs: fi06fl2.x1
          Contact: S.L. Johnson
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: est@watson.wustl.edu
          Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
          Sequencing by: Washington University Genome Sequencing Center
          Seq primer: T3 ET from Amersham
          High quality sequence stop: 521.
FEATURES Location/Qualifiers
    source   1..614
              /organism="Danio rerio"
              /strain="AB"
              /db_xref="taxon:7955"
              /clone="2600399"
              /clone_lib="Sugano Kawakami zebrafish DRA"
              /sex="mixed (one male and one female, including
              unfertilized eggs)"
              /dev_stage="adult"
              /lab_host="DH10B (phage resistant)"
              /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTG);
              Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
              with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTTT];
              double-stranded cDNA was ligated to a DraIII adaptor
              [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
              sites of the pME18S-FL3 vector (5' site CACTGTG, 3' site
              CACCATGTG). XhoI should be used to isolate the cDNA
              insert. Size selection was performed to exclude fragments
              <1.5kb. Library constructed by Dr. Sumio Sugano
              (University of Tokyo Institute of Medical Science) and
              kindly donated by Dr. Koichi Kawakami. Custom primers for
              sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
              primer CGACCTGCAGCTCGAGCACA. "

BASE COUNT   164 a   153 c   144 g   153 t
ORIGIN

Query Match   41.8%; Score 33.4; DB 111; Length 614;
Best Local Similarity 65.3%; Pred. No. 0.66;
Matches   49; Conservative   0; Mismatches   26; Indels   0; Gaps   0;

Qy   4 cacatccaccagccatgagaggtgcttgtactagctctgtgtggctctcgcagtggggg 63
          | |-|||||||||||| |||||| |||| || ||-|| || |-|||| || |||| ||||
Db   1 CAAAACCACCAGCCATGAGAGCTGTGTGCCTTGCCTTGACCTGACGTGAGCCCTGGCATGTC 60

Qy   64 accagtccaacttgg 78
          ||-| | ||||| ||
Db   61 AACAATTCAACCTTG 75

RESULT 14
BF157419
LOCUS   BF157419   740 bp   mRNA   EST   30-OCT-2000
DEFINITION f138b06.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            3816130 5' similar to TR:O93605 093605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   BF157419
VERSION   BF157419.1 GI:11052544
KEYWORDS   EST.
SOURCE   zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 740)
AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy,
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,

---

TITLE   WashU Zebrafish EST Project 1999
JOURNAL   Unpublished (1999)
COMMENT   Other_ESTs: fi20c02.x1
          Contact: S.L. Johnson
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: est@watson.wustl.edu
          Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
          Sequencing by: Washington University Genome Sequencing Center
          Seq primer: T3 ET from Amersham
          High quality sequence stop: 519.
FEATURES Location/Qualifiers
    source   1..764
              /organism="Danio rerio"
              /strain="AB"
              /db_xref="taxon:7955"
              /clone="2601698"
              /clone_lib="Sugano Kawakami zebrafish DRA"
              /sex="mixed (one male and one female, including
              unfertilized eggs)"
              /dev_stage="adult"
              /lab_host="DH10B (phage resistant)"
              /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTG);
              Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
              with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTTTT];
              double-stranded cDNA was ligated to a DraIII adaptor
              [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
              sites of the pME18S-FL3 vector (5' site CACTGTG, 3' site
              CACCATGTG). XhoI should be used to isolate the cDNA
              insert. Size selection was performed to exclude fragments
              <1.5kb. Library constructed by Dr. Sumio Sugano
              (University of Tokyo Institute of Medical Science) and
              kindly donated by Dr. Koichi Kawakami. Custom primers for
              sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
              primer CGACCTGCAGCTCGAGCACA. "

BASE COUNT   204 a   192 c   180 g   188 t
ORIGIN

Query Match   42.0%; Score 33.6; DB 111; Length 764;
Best Local Similarity 66.7%; Pred. No. 0.6;
Matches   48; Conservative   0; Mismatches   24; Indels   0; Gaps   0;

Qy   7 atcccaccagccatgagggtgcttgtactagctctgtgtggctctcgcagtggggacc 66
          | ||||||||||||| |-|||| ||-|| || || || ||-|||| |||| |- || |
Db   1 AACCACCAGCCATGAGAGCTGTGTGCCTTGCCTTGACCTGACGTGAGCCCTGGCAGTCAAC 60

Qy   67 agtccaacttgg 78
          || | ||||| |
Db   61 AGATGAACCTTG 72

RESULT 13
AW128094
LOCUS   AW128094   614 bp   mRNA   EST   25-OCT-1999
DEFINITION fi06f12.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            2600399 5' similar to TR:O93605 093605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   AW128094
VERSION   AW128094.1 GI:6115998
KEYWORDS   EST.
SOURCE   zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 614)
AUTHORS   Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter
            ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE   Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
         and Wilson,R.
JOURNAL WashU Zebrafish EST Project 1998
         Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
         Washington University School of Medicine
         4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
         Tel: 314 286 1800
         Fax: 314 286 1810
         Email: zbrafish@watson.wustl.edu
         Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
         Sequencing by: Washington University Genome Sequencing Center Clone
         distribution information can be found through the I.M.A.G.E.
         Consortium/LLNL, send email to: info@image.llnl.gov
         Seq primer: T3 ET from Amersham
         High quality sequence stop: 509.
FEATURES   Location/Qualifiers
     source   1..740
                 /organism="Danio rerio"
                 /strain="AB"
                 /db_xref="taxon:7955"
                 /clone="3816130"
                 /clone_lib="Sugano Kawakami zebrafish DRA"
                 /sex="mixed (one male and one female, including
                 unfertilized eggs)"
                 /dev_stage="adult"
                 /lab_host="DH10B (phage resistant)"
                 /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                 Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                 with an oligo(dT) primer [ATGTGGCCCTTTTTTTTTTTTTTTTT];
                 double-stranded cDNA was ligated to a DraIII adaptor
                 [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                 sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                 CACCATGTG). XhoI should be used to isolate the cDNA
                 insert. Size selection was performed to exclude fragments
                 <1.5kb. Library constructed by Dr. Sumio Sugano
                 (University of Tokyo Institute of Medical Science) and
                 kindly donated by Dr. Koichi Kawakami. Custom primers for
                 sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                 primer CGACCTGCAGCTCGAGCACA. "
BASE COUNT   196 a   180 c   174 g   189 t   1 others
ORIGIN

Query Match   40.8%; Score 32.6; DB 144; Length 740;
Best Local Similarity 66.2%;   Pred. No. 1.3;
Matches   47; Conservative   0; Mismatches   24; Indels   0; Gaps   0;

Qy   8 tccaccagccatgagggtgcttgtactagctcttgtgtggctctcgcagtggggacca 67
          || ||||||||||||||| | |||| || | |||| |||| ||||| |||| | | ||
Db   35 TCCACCAGCCATGAGAGCTGTTGTGCCTTGCCTTGACTGTAGCCCTCGTGGCATGTCAACA 94

Qy   68 gtccaacttgg 78
          | |||| || |
Db   95 ATTCAACCTTG 105

RESULT 15
AW343221
LOCUS   AW343221   667 bp   mRNA   EST   31-JAN-2000
DEFINITION fi74g10.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            2643426 5' similar to TR:O93605 O93605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION   AW343221
VERSION   AW343221.1 GI:6839587
KEYWORDS   EST.
SOURCE   zebrafish.
  ORGANISM Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 667)
  AUTHORS   Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,